RMHIDD: A Reddit Mental Health Intervention Dialogue Dataset

Anonymous EMNLP submission

Abstract

In the modern human society, mental health is one of the most critical concern. Over past many years a large proportion of population has been affected with serious mental disorders. People with mental illness require effective mental health intervention and treatment as early as possible to decrease the chances of any further mental defilement. In this paper, we present RMHIDD, a new dialogue corpus for automated mental health intervention. The dataset is consists of over 200K Reddit posts collected from 18 different sub-Reddit groups with each post consisting of sequential conversation between the users. On this dataset, we also trained various models for dialogue generation task, namely-'Seq2Seq', 'BART' and 'DialoGPT'. In our analysis we found that the BART model outperformed other models with a higher Perplexity score of 19.7. We also found that the DialoGPT model surpasses other models on various machine translation evaluation metrics. The results generated from various language models were promising and showed the possibility of building automated mental health intervention.

1 Introduction

Mental health is one of the most serious global concerns. In the last few years, there has been a huge increase in the number of people affected by some kind of mental disorder. A report from World Health Organization (WHO)[1] states that 1 out of every 4 people in the world is affected by some kind of mental disorder in their different stages of life. According to WHO's report on depression, it indicates that around 322 million people all around the globe have been affected by depression and this accounts for 18% growth in the total cases from 2005 to 2015. Other major mental illnesses such as anxiety and bipolar disorder which has affected around 264 million people and 60 million people worldwide respectively. Despite increased awareness about mental health conditions and it's management, a report from WHO shows that in every 4 people 3 of them who are suffering from serious mental illness lack timely treatment which pushes them into a major serious mental disorder state. This is a fearful condition and in many cases due to the lack of timely medical help people with mental disorders tend to commit suicide. Every year around 800 thousand people die because of suicide(Organization et al., 2017). Mental health systems are underdeveloped and are not sufficient enough to reach every person who's in need. The basic measures for the prevention of mental illness are psychological intervention and oral consultation. Due to insufficient medical facilities(Jacob and Patel, 2014), the majority of people remain deprived of much-needed treatment and support. Approximately 45% of total world population is living in countries where for every 100K people lesser than 1 psychiatrist is available [2]. Moreover, a report from WHO states that approximately 76% to 85% of people having a mental illness and living in countries with medium and low income do not get the necessary treatment. For high-income countries, it ranges between 35% to 50%. A combination of various factors such as social stigma (Barney et al., 2006) against mentally ill people in the society, unwillingness, or hesitation in asking for help/support, resource scarcity is few reasons behind mismanagement of mental health conditions. Additionally, For the past few years due to the rise in popularity of social media platforms, millions of people are using these platforms to either provide or receive mental health support. Through these online mediums, people express their feelings more

[1] Mental health action plan 2013 - 2020, available at https://www.who.int/mental_health/

[2] Available at http://www.who.int/mental health/evidence/atlas

freely and seek help without any hesitation. One of the most popular online social media platform for sharing mental disorder experiences is Reddit.

Reddit is consists of various subject-specific communities called 'subreddits' where people post their thoughts (topic) and other users can reply through commenting on the post. In this paper, we collected[3] over 200K human-generated posts from various mental disorders help subreddits in a nested way in order to preserve the sequence of conversations (dialogues) generated between the users on a Reddit post. The major **motivation** behind creating and publishing the RMHIDD dialogue corpus is to enable researchers and scientist all around the world to utilize latest advancements in natural language processing and understanding and develop innovative automated mental health intervention tools (such as intelligent chatbots, e-therapy, e-screening, detecting and predicting mental disorders through dialogues) for addressing and solving the issue of mental disorder in our society. We also performed the task of automated dialogue generation that involves generating helpful/supportive dialogues based on the user's mental illness. We used various language models namely - 'Seq2Seq(Sutskever et al., 2014)', 'BART(Lewis et al., 2019) 'and 'DialoGPT(Zhang et al., 2019)' and compared their performance on our proposed dataset. In our experiment, we trained the Seq2Seq model whereas the weights of BART and DialoGPT models were fine-tuned on our dataset. We also performed a comparative study of all the models by evaluating them using various automatic evaluation metrics. The responses generated by the dialogue generation models were very promising and demonstrated the potential and application of natural language processing in the field of automated healthcare systems.

The rest of the paper is organized in the following way. In the section 2 we have summarised the related work done in the field of analyzing user-generated data for mental disorders. In Section 3 we discuss the data collection method and involved steps and in section 4, we discuss various language models used in experiment. Section 5 contains the sequential experiment setup. Section 6 presents analysis and discuss of experiment results and finally in section 7 we conclude the paper.

[3]We will make the data publicly available

2 Related Work

A lot of works have been done that is based on collecting user-generated online data and utilizing it for the analysis and creating insights into mental disorder in people. The author in (De Choudhury and De, 2014) presented a study on characteristics shown by patients with mental illness on social media platforms (Reddit) such as self-disclosure, anonymity, and how it affects the social support received by the patients. The author in papers (Gkotsis et al., 2016), (Park and Conway, 2018) investigated the linguistic characteristic specifically present in the Reddit posts concerned with mental health and illness. In this paper(Gkotsis et al., 2017) the author performed the analysis of Reddit posts and proposed a deep learning-based detection and classification Reddit posts in 11 fine-grained classes of mental disorder. Another paper (Thorstad and Wolff, 2019) also presented an automated mental disorder detection model trained using clinical subreddits which focuses on lexical features in user-generated data to detect the mental illness present in the user. This work presented an automated system for targeted mental health intervention based on user-generated data. In the paper (Shen and Rudzicz, 2017) the author builds a dataset consisting of anxiety-related user-generated posts. The author also applied topic analysis, vector embeddings, emotional norms, and N-gram language modeling for generating features to classify posts in anxiety levels. In paper (Abd Yusof et al., 2018), the author developed lexical features for depression classification tasks and created a dataset using LiveJournalhttp://www.livejournal.com to evaluate feature effectiveness. In the paper (Wongkoblap et al., 2018), the author investigated the relationship between depression and life satisfaction using Facebook user's data and also presented a multilevel predictive model for finding depression in users. Number of researches focusing on identifying depression, anxiety, suicide and bipolarity in social media networks has been done such as (Murrieta et al., 2018), (Lee et al., 2018), (Chen et al., 2018), (Leis et al., 2019), (Wolohan et al., 2018), (Wongkoblap et al., 2019), (Gruda and Hasan, 2019), (Sahota and Sankar, 2020), (Baba et al., 2019). Lot of work focusing on automated healthcare facilities has been done(Liliana Laranjo, 2018). In the papers (Lucas et al., 2017), (Philip et al., 2017), (Tanaka et al., 2017), authors used sequence based step-by-step guided conversation models. Recently neural network based medical dialogue generation models were also proposed. In paper (Wei et al., 2018), author utilized reinforcement learning and developed task oriented dialogue system for automatic medical diagnosis. Paper (Xu et al., 2019) proposed a knowledge-routed relational dialogue system.

Despite plenty of existing research work and resources available on mental disorders, the count of people affected with mental illness rises sharply each year. Most of these works attempt to comprehend the user's action and behavior over social media platforms and develop methods/models to detect the degree of mental disorder among the people. In this paper, we focused on creating a dataset that comprises instances of dialogues between mental disorder help seekers and support providers, collected from social media platforms where people are free to express themselves. Through our work, we want to take a step forward towards developing automated mental health intervention systems that would be readily available to the people suffering from mental disorders.

3 Dataset

The prime source for the collection of our data was Reddit [4]. Reddit is basically a social media website comprised of multiple distinct online communities called subreddits. These are topic-specific forums dedicated to a single topic(e.g., depression, relationship advice, anxiety, etc.) where a user creates a topic, expresses themselves and other users can comment or vote for other comments on that post. We scrapped posts from various mental health management, advice, or support providing subreddits.

- **Mental disorder subreddits** : `r/depression`, `r/anxiety`, `r/stress`, `r/BipolarReddit`
- **Advice Support subreddits** : `r/therapy`, `r/depression_help`, `r/Anxietyhelp`, `r/SuicideWatch`, `r/relationship_advicer`, `r/offmychest`, `r/askatherapist`, `r/relationships`
- **Motivating Uplifting subreddits** : `r/TheMixedNuts`, `r/MadeMeSmile`, `r/FreeCompliments`, `r/UpliftingNews`, `r/DecidingToBeBetter` `r/GetMotivated`,

In the table 1, we have described the number of subreddits used for collecting posts along with the number of dialogues, utterances and the average token length per dialogue. We collected all the content posted in the duration of one year (2, July 2019 to 2, July 2020) on the above-mentioned subreddits. Using `PRAW` API [5] to extract all the content in a nested manner to conserve the sequence of dialogues(topic and comments) in the post. While scrapping dialogues from the posts we removed unwanted bot auto-generated texts by skipping those lines. We also normalized the scrapped dialogues by removing the curse words. We also removed the posts with no comments, in total, we extracted around 200K online user-generated Reddit posts containing instances of dialogues between the users. Table 2 presents the extracted dialogue example from our collected dataset.

4 Methods

In this section, we gave an overview of various state-of-the-art and well-established dialogue generation models. For our experiment, we used 3 deep learning encoder-decoder based models, i.e., Seq2Seq(Wu et al., 2016), DialoGPT(Zhang et al., 2019) and BART(Lewis et al., 2019). We trained the Seq2Seq model on our dataset, whereas for DialoGPT and BART we fine-tuned these models on our dataset.

For a given dialogue having an alternating sequence of utterances between the users, we decided to take two utterances, (i) D_1 (person A issue) the content of the main topic created by a user and (ii) D_2 (person B response) the comment with the highest number of upvotes. So, for each dialogue, we created a pair of utterances, i.e. $\{D_1, D_2\}$ which is used for training all of our dialogue generation models. Given an input D_1, the dialogue generation model outputs D_2.

4.1 Seq2Seq

We utilized an encoder-decoder framework for dialogue generation tasks. Following the original architecture proposed by the author(Wu et al., 2016) for machine translation tasks, we build an LSTMs based deep seq2seq model with attention. The

[4]Available at https://old.reddit.com/

[5]https://github.com/praw-dev/praw

Subreddit Group	#Subreddits	#Dialogue	#Utterances	Avg Tokens/Dialogue
Mental disorder subreddits	4	100,492	246,077	187
Advice Support subreddits	8	83,065	172,688	191
Motivating Uplifting subreddits	6	41,533	95,102	183
Total	18	225,090	513,867	516

Table 1: Data Statistics: We listed 3 subreddit group along with the their associated total number of dialogues, utterances and average tokens per dialogue present in the dataset.

model takes D_1 as input and outputs D_2 as the generated dialogue. Each of the encoder and decoder was consist of 2 LSTM and 1 BiLSTM layer. The input was first passed through 2 LSTM layers, followed by a single BiLSTM layer which generated the latent representation. Similarly, in the decoder, we applied the same 2 LSTM layers with the final BiLSTM layer as the decoding layer.

For a given training set S, we intend to make the log probability of the output sequences T maximum where the given input sequences S given(Sutskever et al., 2014).

$$\frac{1}{|S|}\sum_{(T,S)\epsilon S} log_p(T|S) \tag{1}$$

Once the training is done, according to the LSTM the most probable output sequence is produced:

$$T` = \underset{\text{T}}{\text{argmax}} p(T|S) \tag{2}$$

To obtain final predictions, in the decoder we used the softmax layer and performed decoding using beam search [6]. Finally, the obtained outputs were passed into the loss function, and parameters were updated through backpropagation. Adam(Kingma and Ba, 2014) optimizer was used in the model.

4.2 DialoGPT

In the paperrad2m018iproving, the author proposed a transformer based language model- GPT. For a given token sequence $x_1, ..., x_n$, in a language model the probability over sequence was defined as: $p(x_1, ..., x_n) = \Pi_{i=2}^{n} p(x_i|x_1, ..., x_{i-1})$, where historical sequences are used for predicting the next token. In case of GPT, transformer decoder was used to define $p(x_i|x_1, ..., x_{i-1})$. The decoder consists of stacked self-attention feed-forward layers(each accompanied by normalization layer) for encoding $x_1, ..., x_{i-1}$ and which was then used to predict x_i. In the case of GPT-2(Radford et al., 2019) which was an improvement over GPT, the normalization layer was moved to each of the subblocks input. An extra normalization layer was added after the last self-attention block.

For our experiment, we used DialoGPT(Zhang et al., 2019) which was a GPT-2 based model trained on a very large corpus consisting of English Reddit dialogues. The corpus was consist of 147,116,725 instances of dialogues, collected over a period of 12 years. The model takes the dialogue utterances history S and ground truth response $T = x_1, ..., x_n$, the DialoGPT model aims at maximizinging the probability:$p(T|S) = p(x_1|S)\Pi_{i=2}^{n} p(x_i|S, x_1, ..., x_{i-1})$, where the transformer model defines the conditional probabilities. Through a maximum mutual information (MMI) function(Li et al., 2015), the model also gets penalized for generating uninteresting responses. In our experiment, we used $DialoGPT_{small}$ with 117 million weight parameters.

4.3 BART

BART(Lewis et al., 2019) is a denoising autoencoder that tries to rebuild a corrupted document by performing masked token prediction with the help of bidirectional encoding methods and generates text regressively for natural language generation tasks using a masked attention mechanism. The mask attention mechanism enables the BART model to train on sequence from left to right, generating texts based on the left part of the sequence.

For this transformer-based dialogue system, we create a BART language model wrapper which includes the API of the BART-large model from hugging face-transformers. This pretrained model has 400M trainable parameters with 6 encoding and decoding layers in each block, 16 attention heads both at the encoding and decoding layer. We represent each encoder layer as an $Encoder(.)$ which outputs the hidden state of the respective layer. We

[6] https://google.github.io/seq2seq/nmt/decoding-with-beam-search

Dialogue
creator_id : I am , stuck .
creator_id : I deal with social anxiety , and lately things have been worse than ever and I don ' t know what I should do .
creator_id : To give you an idea with what I've been feeling ; I've always disliked myself, but lately it's become a real hatred .
creator_id : I see any little thing about myself and I feel disgusted , and angry , I can't even take photos of myself or feel comfortable when others do because I know when I see it I ' ll feel repulsed ..
creator_id : And whenever my friends try to make plans I feel unmotivated , and afraid , and I usually make up some stupid lie to get out of things .
creator_id : And I can ' t make plans because I don ' t want to come off as clingy or whatever, and it's really frustrating .
creator_id : Things aren't getting better, but they're not getting worse, it's like this, numbing pain been going on for so long it's frustrating and I'm sick of it .
creator_id : If you know what I'm talking about, if you've felt this before, tell me what to do next.
Commenter_id(1) : I have felt literally the exact same type of way you are describing .
Commenter_id(1) : One thing I would suggest is to stop preparing for the anxiety to come .
Commenter_id(1) : Sometimes we have a tendency to constantly prepare for " war " which keeps us in this exhausting loop of hypervigilance .
Commenter_id(1) : Here is an article that has really helped me to stop letting my triggers have control of when my anxiety pops up. Commenter_id(1) : Please read https://www.thatanxietyguy.com /
Commenter_id(2) : Keep a journal ... write down everything you feel, describe it in as much detail as you can .
Commenter_id(2) : If you have trauma in your past , write about it , try to make written connections between what you feel now and other events in your past where you felt the same .
Commenter_id(2) : For me , my anxiety was due to past abuse , so as an adult I became an approval seeker , validation seeker and people pleaser in an attempt to gain certainty, safety and self esteem from mjy environment .
Commenter_id(2) : It's a hard road , but you need to cross the bridge of "I Don't Give a Crap"... easier said than done but you need to realize that you don't need validation from others, you give it to yourself, give yourself permission for everything you do think or say, you don't need to control whether people and pleased with you, you don't owe anyone an explanation for being who you are, feeling what you feel, wanting what you want... the hatred you feel for yourself could possibly be linked to how you perceived a parent felt about you or treated you when you were younger.
Commenter_id(2) : Please read my other posts ... they may be helpful.

Table 2: An example dialogue in the RMHIDD dataset. creator_id represents the username of the post(issue) creator, Commenter_id represents the commented username. Here on one post have two comments

feed the encoder block of the BART model with a set of input id's from the dialogue history Q. Let the input for the first encoder layer be h_e^0. The h_e^0 is converted into an embedding matrix which passes through the 1^st layer's encoder function yielding a hidden state for the first layer. This step is repeated for each l^{th} layer, where $l \in \{1, .., 12\}$. We get a hidden state h_e for every l^{th} layer by applying the $Encoder(.)$ function as shown in equation(3). The final 12^{th} layer of encoder block output its hidden state h_e^12 which is utilized by the hidden state of the decoder layer for sequential decoding.

$$h_e^l = Encoder(h_e^{l-1}) \tag{3}$$

$$h_d^l = Decoder(h_d^{l-1} \cdot h_e^{12}) \tag{4}$$

Next we feed the set of target-response $T = \{x_1, x_2, ..., x_n\}$ to the decoder block. Similar to encoder block we represent each decoding layer as the $Decoder(.)$ function, which generates hidden state h_d for each decoder layer. Again, let the decoder's input be h_d^0. We feed the model's decoder block with decoder's input ids along with the hidden state of 12^{th} encoder layer. With the help of the decoder block function it generates the hidden state h_d for each layers as shown in equation(4).

In the BART's language model wrapper we have included a linear layer, which generates output to-

Creator_id: what is wrong with me, and should i seek help ? in the 22 years i've been alive, i've had depression for most of it for various reasons. i'm currently in my last year of college and currently have the worst bout i've ever experienced. i'm very close to finishing, yet i find it hard to focus on schoolwork because i often have my mind clouded by ideas that none of this is worth it, and that i shouldn't bother trying . as a result, it's next to impossible to focus on homework which leads to procrastination, and class time is hard to engage in. i find myself unmotivated and feel trapped in a spiral that will lead to inevitable failure. this is a problem that has persisted throughout my college career but has hit me harder now more than ever. i know these thoughts are not true, but it still affects me nonetheless .i often end up stressing because i keep shirking my work and thus continue to put it off. i don't believe i'm on a path to self- destruction, but i don't want this problem to affect my life once i graduate . is this some problem i need addressed or is it just me being lazy? what should i do ?
Ground truth Response: it is totally fine to have some concern about your future since you are about to finish college. you should see a therapist as most colleges offer therapy for free in college counseling centers. you've essentially already paid for it as part of your tuition fees. you should relax and take some break from the college. finding new hobbies, making new friends can help you. in your case, it is best to talk to someone who is an expert in this field. all the best for the future.
Seq2Seq: please be careful and happy. you should take a break and therapy is good for you. go to a doctor. all the best.
DialoGPT: sorry to hear about it. sleep little bit also do exercise daily. if you feel sick you should go to a doctor. seek for support from expert. it is all right. please read my article. thanks
BART: i can understand that college is difficult. just be confident i suggest you to take the therapist help and get professional help from a family. make new friends and talk to your friends. college counselors are good for you. don't give up all the best.

Table 3: Generated responses from various models on a test dialogue

kens probabilities(logits) by applying a normalized exponential function(softmax). This output helps in determining the words within a sequence. Our fine-tuned model aims at maximizing the likelihood as stated in equation(5) by training θ parameters on minimizing cross-entropy of BART model as stated in equation(6)

$$P(T|Q) = P(x_1|Q) \prod_{i=2}^{n} P(x_i|Q, x_1, ..., x_{i-1}) \tag{5}$$

$$\mathcal{L}_{xe}(\theta) = -logP_\theta(T|Q) = -\sum_{t=1}^{N} logP_\theta(y_t|y_{t:t-1}, Q) \tag{6}$$

5 Experiment

In this section, we elaborated on the data preprocessing steps involved in structuring the dataset. We also discussed the hyperparameter setting and optimization strategies used for training the dialogue generation model.

5.1 Data Preprocessing

For our experiment, due to computational limitations, we trained and evaluated all of our models on a randomly collected subset dataset with the size of 50,000 dialogues. As described above4, each instance of dialogue in the dataset we structured them into a pair of utterances('issue' and response'). As shown in the table 3, *Creator_id* (input utterance) and *Ground truth Response* (ground-truth response utterance) were used for training the models. For a given pair of utterance $\{D_1, D_2\}$ in the dataset, we removed all the emojis, unnecessary symbols, and characters. We also replaced the most common abbreviations of words with their original form. We corrected the words with the misspelling. All the unwanted extra spaces in the utterances were removed. For the experiment we divided the dataset into 3 parts: train/validate/train, The distribution of dataset across data was 70%/20%/10% respectively. Hyperparameters were fine-tuned using the validation data.

5.2 Experiment Settings

Seq2Seq : In our Seq2Seq model, we used an embedding layer (trainable matrix) with dimension size of 128. The 2 LSTM layer present in the encoder and decoder was consist of 128 cells. Each of the forwarding and backward LSTM cell in the

	Seq2Seq	**DialoGPT**	**BART**
Perplexity	225.3	27.2	**19.7**
NIST-4	0.60	**1.82**	1.56
BLEU-2	2.16%	**9.19%**	7.38%
BLEU-4	1.53%	**2.83%**	1.97%
METEOR	3.71%	**8.60%**	7.53%

Table 4: Performance score of various models on automatic evaluation metrics.

single BiLSTM layer also consisted of 128 cells. The training was done using the batch size of 16 with the max input and output sequence length set to 400 and 100 respectively. We trained our model for 64 epochs with an initial learning rate of 0.001. At the dense layer, we applied a dropout with a probability of 0.2, and the beam size was set to 3.

BART : For our experiment, we have used $BART_{base}$. We used the Huggingface transformer $BART_{base}$ model[7] provided by the Facebook. As specified in the paper(Lewis et al., 2019), for BART model we followed the given fine-tuning parameters and train the model for 5 epochs with batch size 64. We use the Adam(Kingma and Ba, 2014) optimizer with the linear warm-up scheduler and an initial learning rate of 4e-5. We fed the encoder with noised input tokens of length 400 with its respective attention mask tokens done by the Byte-pair-encoding tokenizer. Similarly we tokenize decoder input with token length of 100 and feed decoder with its respective attention mask. The model trains both the encoder and decoder architect jointly so we get a score logit from the model. We train the model by calculating the cross entropy loss with label smoothing(factor = 0.1) from the logits. Based on validation score we save the model weights and use it on the test dataset evaluation.

DialoGPT : We used $DialoGPT_{small}$(Zhang et al., 2019) and fine-tuned the model on our dataset. The fine-tuning was done for 5 epochs and the batch size was set as 64. The token length for the encoder and decoder was set to 400 and 100 respectively. Similar to the BART model, we used Adam(Kingma and Ba, 2014) optimizer with along with liner learning schedule. The initial learning rate was decided to be 4e-5. During training cross-entropy loss was calculated with the label smoothing factor of 0.1.

[7]Available at https://huggingface.co/models

6 Results and Discussion

We used five performance evaluation matrix to compare the performance of all dialogue generation models. We calculated the perplexity score, METEOR (Lavie and Agarwal, 2007), BLEU-n (Papineni et al., 2002) score(n = 2 4), NIST-n (Doddington, 2002) score for n = 4. For machine translation NIST, METEOR and BLEU are very frequently used evaluation metrics. They compute the similarity by matching the n-grams ground-truth and the model's generated response. In BLEU n-gram precision is calculated by adding equal weight whereas NIST also calculates the informativeness of a particular n-gram and penalizes the non-informative n-grams. Through Perplexity, we calculated and compared the smoothness and quality of produced responses.

In the table 4, we have summarized the performance result of all the dialogue generation model. The following are the observation we can take from the table. Firstly, the overall performance of pretrained language models was superior to the un-trained Seq2Seq model. The reason behind this was the advantage of transfer learning through which pretrained models effectively leverages the knowledge extracted from the large data. Secondly, out of all three models, BART achieved the lowest perplexity score of 19.7, whereas $DialoGPT_{small}$ and Seq2Seq achieved a score of 27.2 and 225.3 respectively. The biggest advantage of $BART_{base}$ was that it was trained on much bigger and diverse data in a way to reconstruct the texts from the corrupted documents, which therefore enhanced and increased BART capabilities as compared to other models. Seq2Seq model scored the highest perplexity which was on an average 89% more than the large pretrained models($DialoGPT_{small}$ and $BART_{base}$). The third observation that could be made was on the machine translation benchmark scores such as METEOR, BLEU, and NIST, the best performance was given by the $DialoGPT_{small}$ model. Since the

$DialoGPT_{small}$ model was pretrained on a large Reddit dialogue dataset which gave the model more contextual understanding for handling our dataset and as a result, more related and relevant dialogue n-grams were produced by the model. With the advantage of large pretraining, the $BART_{base}$ model surpassed the Seq2Seq model on all the machine translation benchmark scores. In the table 3, we have provided the generated responses from all the dialogue generation models on an example dialogue from the test dataset. On average the generated dialogues length form, various models was approximately 50.

7 Conclusion and Future Work

In this paper, we have presented a mental health intervention dialogue dataset. We collected a large number of mental disorder related user-generated data from online platform. Using our dataset, we conducted a systematic analysis of various state-of-the-art dialogue generation language models as an attempt to develop automated mental health intervention system . In our study, we discovered that the large pretrained model($DialoGPT_{small}$ and $BART_{base}$) performed better than the un-trained model(Seq2Seq) on the task of dialogue response generation. The results obtained from various models were very promising and shows the potential of developing automated mental health intervention system in future. We believe that this dataset would enable computer scientist to design and develop more sophisticated, intelligent and feasibly available advance mental health intervention systems such as chatbots, that would help millions of people. In future we aim at extending our current work by collecting a large scale user-generated multilingual mental health dialogue dataset. Through this we would be able to develop a multilingual intervention systems that would not be restricted to single language.

References

Noor Fazilla Abd Yusof, Chenghua Lin, and Frank Guerin. 2018. Assessing the effectiveness of affective lexicons for depression classification. In *International Conference on Applications of Natural Language to Information Systems*, pages 65–69. Springer.

Takahiro Baba, Kensuke Baba, and Daisuke Ikeda. 2019. Detecting mental health illness using short comments. In *International Conference on Advanced Information Networking and Applications*, pages 265–271. Springer.

Lisa J Barney, Kathleen M Griffiths, Anthony F Jorm, and Helen Christensen. 2006. Stigma about depression and its impact on help-seeking intentions. *Australian & New Zealand Journal of Psychiatry*, 40(1):51–54.

Xuetong Chen, Martin D Sykora, Thomas W Jackson, and Suzanne Elayan. 2018. What about mood swings: Identifying depression on twitter with temporal measures of emotions. In *Companion Proceedings of the The Web Conference 2018*, pages 1653–1660.

Munmun De Choudhury and Sushovan De. 2014. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *Eighth international AAAI conference on weblogs and social media*.

George Doddington. 2002. Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. In *Proceedings of the second international conference on Human Language Technology Research*, pages 138–145.

George Gkotsis, Anika Oellrich, Tim Hubbard, Richard Dobson, Maria Liakata, Sumithra Velupillai, and Rina Dutta. 2016. The language of mental health problems in social media. In *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*, pages 63–73.

George Gkotsis, Anika Oellrich, Sumithra Velupillai, Maria Liakata, Tim JP Hubbard, Richard JB Dobson, and Rina Dutta. 2017. Characterisation of mental health conditions in social media using informed deep learning. *Scientific reports*, 7:45141.

Dritjon Gruda and Souleiman Hasan. 2019. Feeling anxious? perceiving anxiety in tweets using machine learning. *Computers in Human Behavior*, 98:245–255.

K Stanly Jacob and Vikram Patel. 2014. Classification of mental disorders: a global mental health perspective. *The Lancet*, 383(9926):1433–1435.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Alon Lavie and Abhaya Agarwal. 2007. Meteor: An automatic metric for mt evaluation with high levels of correlation with human judgments. In *Proceedings of the second workshop on statistical machine translation*, pages 228–231.

Kyung Sang Lee, Hyewon Lee, Woojae Myung, Gil-Young Song, Kihwang Lee, Ho Kim, Bernard J Carroll, and Doh Kwan Kim. 2018. Advanced daily prediction model for national suicide numbers with social media data. *Psychiatry investigation*, 15(4):344.

Angela Leis, Francesco Ronzano, Miguel A Mayer, Laura I Furlong, and Ferran Sanz. 2019. Detecting signs of depression in tweets in spanish: behavioral and linguistic analysis. *Journal of medical Internet research*, 21(6):e14199.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*.

Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. 2015. A diversity-promoting objective function for neural conversation models. *arXiv preprint arXiv:1510.03055*.

Huong Ly Tong Ahmet Baki Kocaballi Jessica Chen Rabia Bashir Didi Surian Blanca Gallego Farah Magrabi Annie YS Lau Liliana Laranjo, Adam G Dunn. 2018. Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association*.

Gale M Lucas, Albert Rizzo, Jonathan Gratch, Stefan Scherer, Giota Stratou, Jill Boberg, and Louis-Philippe Morency. 2017. Reporting mental health symptoms: breaking down barriers to care with virtual human interviewers. *Frontiers in Robotics and AI*, 4:51.

Julissa Murrieta, Christopher C Frye, Linda Sun, Linh G Ly, Courtney S Cochancela, and Elizabeth V Eikey. 2018. # depression: Findings from a literature review of 10 years of social media and depression research. In *International Conference on Information*, pages 47–56. Springer.

World Health Organization et al. 2017. Depression and other common mental disorders: global health estimates. Technical report, World Health Organization.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.

Albert Park and Mike Conway. 2018. Harnessing reddit to understand the written-communication challenges experienced by individuals with mental health disorders: Analysis of texts from mental health communities. *Journal of medical Internet research*, 20(4):e121.

Pierre Philip, Jean-Arthur Micoulaud-Franchi, Patricia Sagaspe, Etienne De Sevin, Jérôme Olive, Stéphanie Bioulac, and Alain Sauteraud. 2017. Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders. *Scientific reports*, 7(1):1–7.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners. *OpenAI Blog*, 1(8):9.

Puneet KC Sahota and Pamela L Sankar. 2020. Bipolar disorder, genetic risk, and reproductive decision-making: A qualitative study of social media discussion boards. *Qualitative health research*, 30(2):293–302.

Judy Hanwen Shen and Frank Rudzicz. 2017. Detecting anxiety through reddit. In *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology—From Linguistic Signal to Clinical Reality*, pages 58–65.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112.

Hiroki Tanaka, Hideki Negoro, Hidemi Iwasaka, and Satoshi Nakamura. 2017. Embodied conversational agents for multimodal automated social skills training in people with autism spectrum disorders. *PloS one*, 12(8):e0182151.

Robert Thorstad and Phillip Wolff. 2019. Predicting future mental illness from social media: A big-data approach. *Behavior research methods*, 51(4):1586–1600.

Zhongyu Wei, Qianlong Liu, Baolin Peng, Huaixiao Tou, Ting Chen, Xuan-Jing Huang, Kam-Fai Wong, and Xiang Dai. 2018. Task-oriented dialogue system for automatic diagnosis. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 201–207.

JT Wolohan, Misato Hiraga, Atreyee Mukherjee, Zeeshan Ali Sayyed, and Matthew Millard. 2018. Detecting linguistic traces of depression in topic-restricted text: Attending to self-stigmatized depression with nlp. In *Proceedings of the First International Workshop on Language Cognition and Computational Models*, pages 11–21.

Akkapon Wongkoblap, Miguel A Vadillo, and Vasa Curcin. 2018. A multilevel predictive model for detecting social network users with depression. In *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, pages 130–135. IEEE.

Akkapon Wongkoblap, Miguel A Vadillo, and Vasa Curcin. 2019. Modeling depression symptoms from social network data through multiple instance learning. *AMIA Summits on Translational Science Proceedings*, 2019:44.

Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, et al. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*.

Lin Xu, Qixian Zhou, Ke Gong, Xiaodan Liang, Jianheng Tang, and Liang Lin. 2019. End-to-end knowledge-routed relational dialogue system for automatic diagnosis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 7346–7353.

Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and Bill Dolan. 2019. Dialogpt: Large-scale generative pre-training for conversational response generation. *arXiv preprint arXiv:1911.00536*.