Generative Adversarial Imitation Learning for Empathy-based AI

Pratyush Muthukumar, [1] **Karishma Muthukumar,** [1]
Deepan Muthirayan, [1] **Pramod Khargonekar,** [1]

[1] University of California, Irvine
Irvine, CA, USA
muthukup@uci.edu, muthukuk@uci.edu,
dmuthira@uci.edu, pramod.khargonekar@uci.edu

Abstract

Generative adversarial imitation learning (GAIL) is a model-free algorithm that has been shown to provide strong results in imitating complex behaviors in high-dimensional environments. In this paper, we utilize the GAIL model for text generation to develop empathy-based context-aware conversational AI. Our model uses an expert trajectory of empathetic prompt-response dialogues which can accurately exhibit the correct empathetic emotion when generating a response. The generator of the GAIL model uses the GPT-2 sequential pre-trained language model trained on 117 million parameters from 40 GB of internet data. We propose a novel application of an approach used in transfer learning to fine tune the GPT-2 model within the generator of the GAIL model to generate concise, user-specific empathetic responses validated against the discriminator within the GAIL model. Our novel GAIL model utilizes a sentiment analysis history-based reinforcement learning approach to empathetically respond to human interactions in a personalized manner. We find that our model's response scores on various human-generated prompts collected from the Facebook Empathetic Dialogues dataset outperform baseline counterparts. Moreover, our model improves upon various history-based conversational AI models developed recently, as our model's performance over a sustained conversation of 3 or more interactions outperform similar conversational AI models.

Introduction

Text generation models designed for intelligent conversation systems are increasingly popular in research. Applications of text generation models in intelligent home systems are also becoming widespread. Many modern text generation models take advantage of deep learning paradigms for natural dialogue generation. State-of-the-art results in this field utilize the numerous processing layers of deep learning architectures typified by Recurrent Neural Networks, Variational Autoencoders, and Transformer models (Zhang et al. 2019; Li et al. 2020; Serban et al. 2017). These neural network approaches are capable of context-aware, history-based text generation, however, it is less often the case we see research focused on developing empathetic dialogue systems. The challenge of developing empathetic dialogue systems is inherently a conceptual one: in literature and application, there is no clear definition of empathy. Cognitive neuroscientists are actively seeking to define empathy empirically, however, even the most objective definitions fail to cater to each individual's perception of empathy (Gerdes, Segal, and Lietz 2010). It is even more challenging to develop artificial empathy capable of detecting and responding to human emotions (Asada 2015; Sharma and Bikshandi 2020).

In this paper, we propose an architecture capable of empathetic, context-aware, natural dialogue generation. To tackle the challenges of such a problem, we use deep reinforcement learning. Deep reinforcement learning has recently been used for traditional text generation, as the nature of reinforcement learning allows for success in complex, high dimensional environments. The challenge of picking an appropriate response out of an entire set of possible utterances in a language is certainly the caliber of problem that reinforcement learning is able to solve. Li et al. (2016) proposed one of the first applications of deep Q-learning for dialogue generation by simulating dialogue between two virtual agents using policy gradient methods. Their work showed improvements over traditional MLE-based Seq2Seq models, citing the main advantages of deep Q-learning to be its ability to keep the conversation moving through non-generic responses and its ability to recognize repetitive utterances.

While traditional reinforcement learning has shown success for natural dialogue generation, a key shortcoming is evident when we introduce empathy-based intelligent conversation agents. For an empathetic dialogue agent to select an optimal policy in a state-space, it must maximize a reward function which defines empathetic conversation. However, the same problem arises: it is virtually impossible to define a statistical formula for empathy.

We propose the application of inverse reinforcement learning or imitation learning to solve this problem in a different way. Instead of postulating an empirical formula for empathy, we feed the model a trajectory of empathetic expert actions that an empathetic dialogue agent can imitate. Our rationale behind this approach is that it is significantly easier to recognize examples of empathy in conversation, actions, or people rather than developing an overarching formula for empathy.

In our implementation, we utilize the generative adversarial imitation learning (GAIL) model developed by Ho and Ermon (2016), a reinforcement learning variant of the generative adversarial network (GAN) developed by Goodfellow et al. (2014), used famously for video and image gen-

eration. We also use the large-scale pre-trained language model GPT-2 within the generator of the GAIL architecture that allows for basic textual understanding and generation (Radford et al. 2019). For the optimization, we utilize proximal policy optimization (PPO) (Schulman et al. 2017). Our GAIL implementation fine tunes the large pre-trained GPT-2 language model within the generator using the discriminator reward signals rather than retraining the language model in the generator.

For a fair evaluation, we utilize an error metric commonly used in the field of natural language processing to evaluate langauge models. We evaluate our model over single-turn and multi-turn conversations using the perplexity and BLEU error metrics (Chen, Beeferman, and Rosenfeld 1998; Papineni et al. 2002). Instead of testing our model's performance over common baselines for natural dialogue generation including COCO, ROCStories, and CommonGEN (Lin et al. 2014, 2019; Mostafazadeh et al. 2017), we evaluate our model against others using our own empathetic dataset. The motivation for such a methodology is that the typical benchmarks used for conditional and unconditional text generation do not contain empathetic prompts or responses, and as a result, our model's effectiveness will not be tested. Moreover, our model does not accomplish the same tasks that would be measured in these text generation baseline datasets.

The novelty of our approach is evident in the way we utilize deep reinforcement learning models for highly accurate, history-based, context-aware, natural empathetic dialogue generation while preserving the complex and adaptive nature of empathy in the generated dialogue. Our model is the first of its kind to solve the unique problem of empathetic dialogue generation using deep reinforcement learning models. Our model's novelty in the method of utilizing deep reinforcement learning serves as a starting point for future research seeking to advance the field of artificial empathy.

We believe that our work makes the following important contributions: (1) we propose an application of the generative adversarial imitation learning architecture for empathetic text generation, (2) we fine tune the large-scale pretrained language model GPT-2 within the GAIL training process using expert empathetic actions to generate empathetic dialogues, (3) we show that our novel architecture outperforms similar text generation models in singe-turn and multi-turn empathetic dialogues using the perplexity and BLEU error metrics, (4) we collect a demonstrative set of empathetic trajectories from open-access data sources for efficient single-turn and multi-turn empathetic dialogue generation.

Related Work

Artificial empathy is a large field, however, research into intelligent conversation agents with empathy is scarce. One of the most famous is a proprietary chatbot systems available to consumers is WoeBot developed by Fitzpatrick, Darcy, and Vierhile (2017). WoeBot is an intelligent conversation agent that allows patients to converse in a structured set of prompts to deal with various mental health issues. The intelligent agent is certified in cognitive behavioral therapy (CBT). Although it does not provide natural context-aware text generation, as the set of prompts are relatively static with respect to the subject of the conversation, WoeBot is able to discover and respond to human emotions.

Fung et al. (2018) is one of the first papers to introduce the application of deep reinforcement learning for empathetic dialogue generation. Their work describes a reinforcement learning variant of a traditional Seq2Seq model used for dialogue generation. Instead of defining a reward function for empathy, they seek to categorize each human prompt into an emotion class. They classify each prompt into an emotion class by the words and emoticons used in each utterance, and then respond using a set of predefined responses based on the discovered emotion class. They train their deep reinforcement learning architecture using Proximal Policy Optimization (PPO).

Wu, Li, and Yu (2020) propose an application of a generative adversarial imitation learning model for traditional text generation. Their model, TextGAIL, shows improvement over previous work on text generation of image captions and baseline text generation tasks including COCO, CommonGEN, and ROCStories. However, since the goal of TextGAIL is text generation rather than dialogue generation, it is not effective at responding to human prompts coherently.

Model Architecture

In this section, we describe the architecture of our proposed model. We describe our novel fine tuning approach on pretrained language models using expert empathetic actions, as well as the structure of the generator and discriminator models within our generative adversarial imitation learning architecture. A visualization of the overall architecture of our model is described in Figure 1.

For dialogue generation, we modify the generative adversarial imitation learning architecture. Generative adversarial imitation learning is a model-free imitation learning architecture able to directly imitate near-optimal expert trajectories in high dimensional, complex environments. The GAIL architecture consists of a generator G and discriminator D, where the generator seeks to generate responses similar to empathetic human responses, while the discriminator seeks to distinguish between the generated responses and the expert empathetic responses and propagate a reward signal back to the generator. As a result, the GAIL architecture does not need to discover the exact reward function that a typical reinforcement learning agent would maximize, which suits our goal of empathetic dialogue generation.

The generator of the GAIL model performs imitation learning by defining the inverse reinforcement learning problem as the dual of the reinforcement learning problem. That is, we define the reinforcement learning problem as

$$RL(c) = \arg\min_{\pi \in \prod} -H(\pi) + \mathbb{E}[c(s,a)], \qquad (1)$$

where $\pi \in \prod$ is a recovered policy, $H(\pi)$ is the $\gamma-$distributed entropy of policy π, and $c(s,a)$ is the cost for state s and action a. Similarly, we then define the ψ-regularized maximum entropy inverse reinforcement learn-

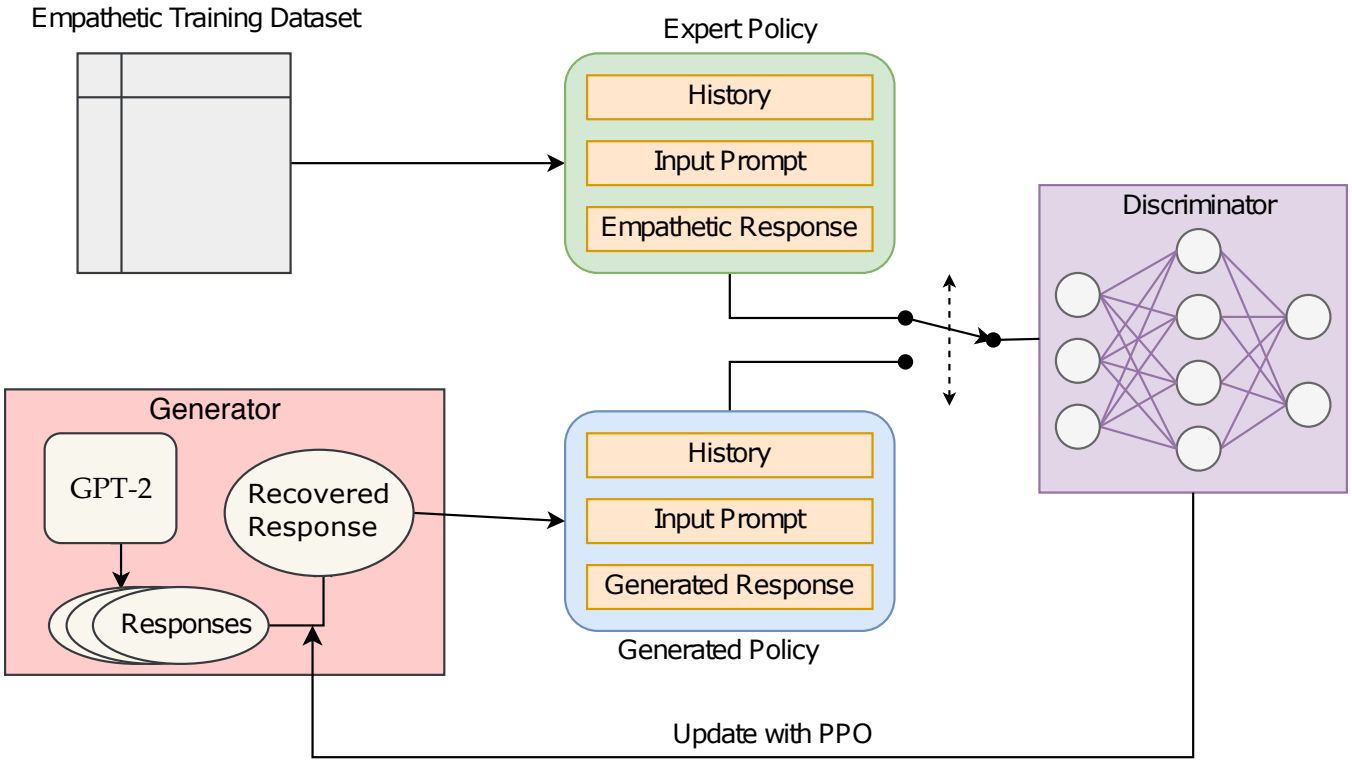

Figure 1: Model Architecture

ing problem as

$$IRL_{\psi}(\pi^E) = \arg\max_{c \in \mathbb{R}^{s \times a}} -\psi(c) + \left(\min_{\pi \in \prod} -H(\pi) + \mathbb{E}_{\pi}[c(s,a)] \right) - \mathbb{E}_{\pi^E}[c(s,a)], \quad (2)$$

where ψ is a regularization term and π^E is the optimal policy. Note that ψ-regularized inverse reinforcement learning implicitly seeks a policy where its occupancy measure ρ is close to the expert.

The generated policy of the GAIL is then

$$RL \circ IRL_{\psi}(\pi^E) = \arg\min_{\pi \in \prod} -H(\pi) + \psi^*(\rho_{\pi} - \rho_{\pi^E}), \quad (3)$$

where ρ_{π} is the occupancy measure of the recovered policy and ρ_{π^E} is the occupancy measure of the expert policy. We can select a regularizer that allows us to implement a more complex class of cost functions using neural networks. In our implementation, we define our cost regularizer as

$$\psi_{GA}(c) := \begin{cases} \mathbb{E}_{\pi^E}[g(c(s,a))] & \text{if } c < 0 \\ +\infty & \text{otherwise} \end{cases} \quad (4)$$

where

$$g(x) = \begin{cases} -x - \log(1 - e^x) & \text{if } x < 0 \\ +\infty & \text{otherwise.} \end{cases} \quad (5)$$

The motivation behind this specific cost regularizer is because the convex conjugate ψ^* is the optimal negative log-loss of the binary classification problem of distinguishing between state-action pairs of π and π_E:

$$\max_{D \in (0,1)^{S \times A}} \mathbb{E}_{\pi}[\log D(s,a)] + \mathbb{E}_{\pi_E}[\log(1 - D(s,a))], \quad (6)$$

which is also exactly the same as the cost function of the discriminator network of a traditional generative adversarial network (GAN).

For dialogue generation, we replace the state s within the GAIL with a two-part input and its corresponding action a with the target response. This two-part input consists of (1) an optional history of the earlier prompts and responses in the conversation and (2) a required input prompt. We also optimize the policy generation within the generator of the GAIL with PPO instead of trust region policy optimization (TRPO) (Schulman et al. 2015), which was used in the original paper. The TRPO algorithm is a constrained optimization problem that ensures that the policy is not moving too far away from the starting point by using the KL-divergence. TRPO is a commonly used policy optimizer, but more recent research has shown newer policy optimizers that have lower variance and do not solely require small steps to convergence (Engstrom et al. 2019). PPO is a simpler, more effective policy optimization algorithm compared to TRPO. PPO tries to compute an update at each time step that minimizes the cost function while ensuring the deviation from the previous policy is small. Instead of a KL-divergence constraint like TRPO, PPO uses a KL-divergence penalty:

$$\max_{\theta} \sum_{n=1}^{N} \frac{\pi_{\theta}(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)} \hat{A}_t - C \cdot KL_{\pi_{\theta_{\text{old}}}}(\pi_{\theta}). \quad (7)$$

For the Generator network, we utilize the pre-trained language model GPT-2, a large-scale multipurpose language model with a transformer architecture developed by OpenAI with 117 million parameters trained on 40 GB of internet data. The nature of our modified GAIL model allows for the expert empathetic actions to fine tune the baseline language model for empathetic dialogue generation in a similar manner to the fine tuning process found in transfer learning. Instead of retraining the entire language model for empathetic dialogues, which would be computationally infeasible due to the size of the language model, the GAIL model first queries the pretrained architecture using the input prompt in multiple batches for each step, treats all possible GPT-2 responses for a prompt as the set of possible actions to take in the state-space, and selects the response most similar with respect to the occupancy score of the optimal expert empathetic response. The GPT-2 model is then fine tuned by updating the self-attention parameters in its transformer architecture. The GPT-2 decoder-only transformer architecture allows us to solely train its high level parameters, the self-attention parameters, by setting the corresponding label of the input prompt as the response with the most similar occupancy score to the expert trajectories (Khandelwal et al. 2019).

We describe the training process of our overall modified generative adversarial imitation learning algorithm in the Appendix (Algorithm 1). We also describe the training process of the generator network within the modified GAIL architecture in the Appendix (Algorithm 2).

Experiments

In this section, we describe the datasets and implementation details of our experiments.

Data

The datasets we used for the following experiments were selected to be clear examples of empathetic conversations, actions, or dialogues. Our dataset plays a larger role in the performance of our model compared to similar research because the empathetic nature of our model comes from imitating our dataset of expert empathetic trajectories. As a result, we chose our dataset carefully in order to exemplify human empathetic actions in a broad setting. All data we use in this paper is open source, anonymized, and does not contain any personally identifiable or offensive content.

The first dataset we used is the EmpatheticDialogues dataset collected by Facebook AI Research (Rashkin et al. 2018). The EmpatheticDialogues dataset consists of 25 thousand single-turn and multi-turn empathetic conversations. Each data sample contains (1) a label which describes the overall emotion of the conversation and (2) the transcript of a conversation between exactly two human speakers. These conversations were human-labeled and human-generated through crowd sourcing via Amazon Mechanical Turk.

The second dataset we used is the DailyDialog dataset collected by Li et al. (2017) consisting of 13 thousand multi-turn empathetic conversations. The data labeling task was crowd sourced and human-labeled. We use this dataset as it provides longer example conversations sustained over a single topic. This allows our model to produce more context-aware and history-based dialogue generation, as it imitates the sustained flow of conversation evident in the DailyDialog dataset.

Finally, we also collect empathetic tweets as a datasource. We find that including empathetic tweets allow our model to offer advice or provide inspirational responses at certain points in the conversation, resulting in a more thoughtful and altruistic conversational agent. We collect the entire Twitter histories of established or generally agreed upon compassionate Twitter accounts including @DalaiLama, @DailyZen, and @MindfulEveryday.

We preprocess data from these three sources into a single dataset of tuples (h_i, p_i, r_i) where h_i is the history, p_i is the prompt, and r_i is the response. h_i is an optional component of the tuple which denotes all utterances prior to the selected prompt and response. p_i is the single utterance human prompt input of the conversation directly preceding the empathetic response. r_i is the single utterance human empathetic response.

To create the expert empathetic trajectories from the dataset, we must pre-process the conversation data. For the EmpatheticDialogues and DailyDialog datasets, we create an expert trajectory for each turn of the conversation, where a turn is defined as a prompt from one speaker and a response utterance from another speaker. If the data sample is a single turn conversation, then we construct an expert empathetic trajectory sample where the history h_i is an empty string, the input prompt p_i is the first speaker's utterance, and the optimal response r_i is the second speaker's utterance. For multi turn conversations, we generate multiple trajectories such that we create a trajectory for the first turn in the conversation with no history h_i, and the first two utterances as prompt p_i and response r_i respectively. For every turn after the first, we concatenate all previous turns before the current turn and store the utterances as the history h_i, the current turn's first speaker utterance as the prompt p_i and the current turn's responding speaker utterance as the response r_i.

For each conversation in either the EmpatheticDialogues or DailyDialog dataset, there is at least one generated empathetic trajectory. In general, there are usually more generated for each data sample, since most conversations in these datasets are multi turn: EmpatheticDialogues has an average of 2.3 turns per conversation; DailyDialog has an average of 4.8 turns per conversation.

For the empathetic tweets that are not in conversation format, we store no history h_i but store the tweet text as both the prompt p_i and the response r_i. That is, $p_i = r_i$ for all tuples constructed from empathetic tweets in the dataset. For each tweet, there is exactly one generated expert empathetic trajectory. In total, we generate 208600 expert empathetic trajectories in the form (h_i, p_i, r_i): 79353 from EmpatheticDialogues, 61935 from DailyDialog, and 67312 from empathetic tweets.

We then partition 146020 expert empathetic trajectories as the training set, 41720 trajectories for the testing set, and 20860 trajectories for the validation set. This partition corre-

Metric	Model	1 Turn	2 Turn	3+ Turn
Perplexity	GPT-2 + MLE	36.61 ± 3.21	29.03 ± 3.02	21.02 ± 2.41
	RL-Seq2Seq	116.96 ± 18.14	86.31 ± 13.23	74.35 ± 10.20
	TextGAIL	13.51 ± 2.39	8.93 ± 1.88	6.90 ± 1.45
	EmpathyGAIL	$\mathbf{7.32 \pm 1.35}$	$\mathbf{5.20 \pm 1.03}$	$\mathbf{4.38 \pm 0.81}$
BLEU	GPT-2 + MLE	5.28 ± 0.05	5.85 ± 0.03	12.55 ± 0.03
	RL-Seq2Seq	1.34 ± 0.01	2.95 ± 0.03	4.13 ± 0.02
	TextGAIL	13.12 ± 0.02	14.28 ± 0.04	20.43 ± 0.05
	EmpathyGAIL	$\mathbf{16.52 \pm 0.06}$	$\mathbf{20.34 \pm 0.15}$	$\mathbf{25.35 \pm 0.12}$

Table 1: Perplexity and BLEU Error Results for 1 Turn, 2 Turn, and 3+ Turn Conversations

sponds to a 70%-20%-10% training-testing-validation split.

Implementation

Our implementation of our modified generative adversarial imitation learning architecture uses a neural network architecture with two hidden layers of 100 units each, with $\tanh$ nonlinearities in between for the generator and discriminator. The pre-trained model weights for the generator network is the base GPT-2 model. A majority of our implementation was completed using the PyTorch Python framework for Python 3.7.7 (PyTorch 2021). All external frameworks used in our implementation including GPT-2 and PyTorch are open-source assets.

We train our model over 750 epochs with a batch size of 32, a validation frequency of 10 epochs, and a human demonstration mix ratio set at 0.3 at the start and decreases linearly. Our generator network optimizes our policy through a custom minimal implementation of PPO with a minibatch size of 8 and an epsilon value of 0.2. We describe additional hyperparameter selections for our model implementation in the Appendix (Table 3).

We train and evaluate our model using the Google Colaboroatory GPU-enabled runtime with a 2496 CUDA core Tesla K80 GPU with 12GB GDDR5 VRAM and a single core multi thread 2.3 gigahertz Intel Xeon CPU. The training process of 750 epochs completed in roughly four compute hours.

Evaluation Metrics

We evaluate our model's performance using the perplexity and BLEU metrics, which are common metrics for evaluating the effectiveness of language models. We compare our model against baseline dialogue generation models. Roughly, perplexity measures the human-readability of a generated text, where a low perplexity score corresponds to an easily understandable output. The perplexity score P for a language model $p_M(\text{next word } w| \text{ history } h)$ on a test set $T = \{w_1, \ldots, w_t\}$ is

$$P(p_M) = \frac{1}{(\prod_{i=1}^{T} p_M(w_i|w_1 \ldots w_{i-1}))^{\frac{1}{t}}}.$$

The Bilingual Evaluation Understudy Score (BLEU) error metric measures the similarity of a candidate sentence to a target sentence. Here, we will measure the BLEU scores on generated responses to optimal empathetic responses in the test set. We use the bigram matching BLEU metric in the range of 0.0 to 100.0, where a high BLEU score denotes a closer similarity between the generated and target utterance.

Results

We test our model against three baselines. For each baseline, due to the nature of our problem, we collect perplexity error by training and evaluating these models on our empathetic dataset as opposed to a common benchmark dataset. For all baselines, we utilize the default hyperparameters described in their respective implementations. All baselines we select are dialogue generation models. Our first baseline is a base GPT-2 dialogue generation instance fine tuned to our dataset through maximum likelihood estimation (MLE). Our second baseline is the reinforcement learning variant of the Seq2Seq architecture developed by Fung et al. (2018) where they classify input prompts into emotion classes to respond empathetically. Our final baseline is the TextGAIL model developed by Wu, Li, and Yu (2020) that utilizes the GAIL architecture for text generation.

We test our model against the baselines using the BLEU and perplexity error metrics for a single turn, 2 turn, and 3+ turn conversations. That is, a single turn conversation contains no history; a 2 turn conversation contains the history of a single turn; and a 3+ turn conversation contains the history of multiple turns. We provide our results in Table 1 along with error bars with respect to various iterations using different random seeds. To display further generalizability of our approach, we also provide BLEU and perplexity scores with error bars of our model against baselines trained on only one of each of the three empathetic data sources (EmpatheticDialogues, Daily Dialog, empathetic tweets) in the Appendix (Tables 4,5,6). We provide a visualization of perplexity and BLEU error over conversation length for each model in Figure 2 and 3 respectively. Our BLEU and perplexity error metric results show that our model, *EmpathyGAIL*, provides significant improvement over baselines.

In order to display the empathetic abilities of our approach, we also provide some example responses from our model. Each example response corresponds to an input prompt and an optional history. Table 2 describes a few of

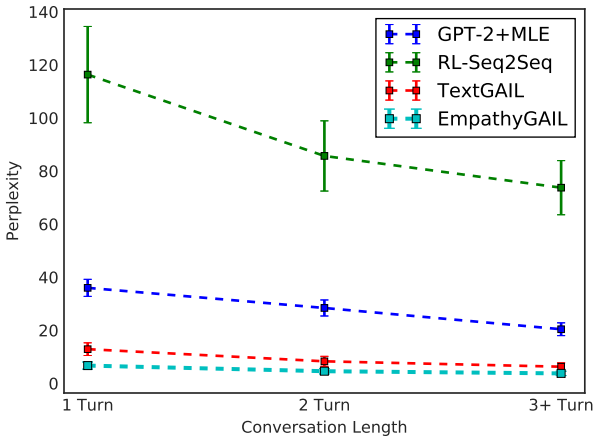

Figure 2: Perplexity scores

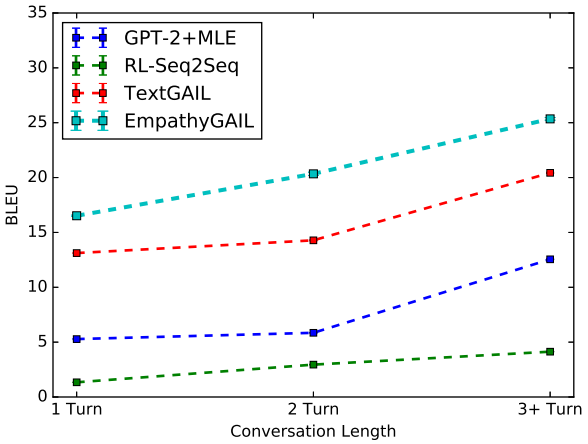

Figure 3: BLEU scores

these qualitative results. We also provide extended qualitative results including the qualitative results of the baseline language models in the Appendix (Table 7).

From the qualitative results of Table 2 it is evident that *EmpathyGAIL* performs at or near the level of ground truth human responses, as validated by the BLEU and perplexity scores on our quantitative results. Furthermore, the qualitative results of the baseline models display the superior empathetic and comprehensible characteristics of dialogue generated from our model.

Conclusion

We propose a generative adversarial imitation learning architecture for empathetic dialogue generation. Our key contributions include a modified GAIL architecture for dialogue generation, a fine tuning approach using expert empathetic trajectories on large-scale pre-trained language models, and our results show improved perplexity and BLEU scores over single turn and multi turn conversations. Our work shows that deep imitation learning models can accurately and effectively provide context-aware, empathetic, and natural dialogue generation.

Our results show an average perplexity error decrease of 80% and a BLEU score increase of 187% compared to a base GPT-2 model fine tuned to our dataset through MLE. Our results also show an average perplexity error decrease of 42% and a BLEU score increase of 31% compared to a generative adversarial imitation learning model for text gen-

	Example Results
History	I moved to LA a few years ago and I never had the chance to visit home.
Prompt	I'm going to see my parents soon!
Generated Response	That's awesome! I hope you have a good time.
Human Response	Great! I'm glad you're happy.
History	None
Prompt	He was a very good friend of mine. Last month, he passed away.
Generated Response	I am very sad to hear that he is no longer alive.
Human Response	I hope you are doing OK.
History	I was in a drive-thru line and someone cut in front of me.
Prompt	I was furious and blew up at him.
Generated Response	Sometimes it makes more sense to just walk away. You cannot control others, you can only control yourself.
Human Response	I find that meditating helps calm me down.
History	None
Prompt	I had a job interview today and I think it went really well.
Generated Response	Wonderful!
Human Response	That's nice to hear!

Table 2: Example Model Results

eration (Wu, Li, and Yu 2020).

There are a few limitations of our model. We find that at certain times, our model will begin to generate fake experiences or stories in an attempt to connect with the input prompt. This is because the input empathetic datasets we use are entirely human generated such that a human responder can empathetically respond to a human prompt by narrating their own prior experiences. Since our architecture seeks to imitate these empathetic responses, we find that our model does the same. Also the GPT-2 pre-trained language model we use within the generator of the GAIL is trained on datasets which contain biases or factual inaccuracies, and as a result will be reflected in the output of our model.

Future Work and Impacts

We hope to use additional sources of empathetic data which are human-labeled but may not be human-generated to ensure that our model does not generate fake experiences or stories to connect with the input prompt. We also hope to implement more powerful pre-trained language models like the recently developed GPT-3 model (Brown et al. 2020). Another direction we would like to take our research is developing a more personalized sense of empathy for each user. Empathy is individual-specific, so we hope to implement a set of calibrating questions that gauge the perception of empathy for a user prior to providing empathetic dialogue. We may then be able to use insights of these calibrating questions as a latent seed for the generator of the GAIL architecture, similar to the technique developed by Chen et al. (2016).

We forsee our impacts as a tool for patients suffering with mental health, anxiety, or depression to converse with an empathetic chatbot at any time. Before making our developments available for public use, we will have to complete rigorous training and testing to ensure that the chatbot provides help and not harm. There may be serious negative societal impacts if adequate testing is not completed, so we will heavily focus our future research into ensuring the safety and effectiveness of our work.

References

Asada, M. 2015. Towards artificial empathy. *International Journal of Social Robotics*, 7(1): 19–33.

Brown, T. B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. 2020. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.

Chen, S. F.; Beeferman, D.; and Rosenfeld, R. 1998. Evaluation metrics for language models.

Chen, X.; Duan, Y.; Houthooft, R.; Schulman, J.; Sutskever, I.; and Abbeel, P. 2016. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 2180–2188.

Engstrom, L.; Ilyas, A.; Santurkar, S.; Tsipras, D.; Janoos, F.; Rudolph, L.; and Madry, A. 2019. Implementation matters in deep rl: A case study on PPO and TRPO. In *International conference on learning representations*.

Fitzpatrick, K. K.; Darcy, A.; and Vierhile, M. 2017. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR mental health*, 4(2): e19.

Fung, P.; Bertero, D.; Xu, P.; Park, J. H.; Wu, C.-S.; and Madotto, A. 2018. Empathetic dialog systems. In *The international conference on language resources and evaluation. European Language Resources Association*.

Gerdes, K. E.; Segal, E. A.; and Lietz, C. A. 2010. Conceptualising and measuring empathy. *British Journal of Social Work*, 40(7): 2326–2343.

Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative adversarial networks. *arXiv preprint arXiv:1406.2661*.

Ho, J.; and Ermon, S. 2016. Generative adversarial imitation learning. *arXiv preprint arXiv:1606.03476*.

Khandelwal, U.; Clark, K.; Jurafsky, D.; and Kaiser, L. 2019. Sample efficient text summarization using a single pre-trained transformer. *arXiv preprint arXiv:1905.08836*.

Li, J.; Monroe, W.; Ritter, A.; Galley, M.; Gao, J.; and Jurafsky, D. 2016. Deep reinforcement learning for dialogue generation. *arXiv preprint arXiv:1606.01541*.

Li, Y.; Feng, R.; Rehg, I.; and Zhang, C. 2020. Transformer-Based Neural Text Generation with Syntactic Guidance. *arXiv preprint arXiv:2010.01737*.

Li, Y.; Su, H.; Shen, X.; Li, W.; Cao, Z.; and Niu, S. 2017. DailyDialog: A manually labelled multi-turn dialogue dataset. *arXiv preprint arXiv:1710.03957*.

Lin, B. Y.; Shen, M.; Xing, Y.; Zhou, P.; and Ren, X. 2019. CommonGEN: A constrained text generation dataset towards generative commonsense reasoning.

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.

Mostafazadeh, N.; Roth, M.; Louis, A.; Chambers, N.; and Allen, J. 2017. Lsdsem 2017 shared task: The story cloze test. In *Proceedings of the 2nd Workshop on Linking Models of Lexical, Sentential and Discourse-level Semantics*, 46–51.

Papineni, K.; Roukos, S.; Ward, T.; and Zhu, W.-J. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, 311–318.

PyTorch. 2021. PyTorch. https://pytorch.org/.

Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; and Sutskever, I. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8): 9.

Rashkin, H.; Smith, E. M.; Li, M.; and Boureau, Y.-L. 2018. Towards empathetic open-domain conversation models: A new benchmark and dataset. *arXiv preprint arXiv:1811.00207*.

Schulman, J.; Levine, S.; Abbeel, P.; Jordan, M.; and Moritz, P. 2015. Trust region policy optimization. In *International conference on machine learning*, 1889–1897. PMLR.

Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; and Klimov, O. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

Serban, I.; Klinger, T.; Tesauro, G.; Talamadupula, K.; Zhou, B.; Bengio, Y.; and Courville, A. 2017. Multiresolution recurrent neural networks: An application to dialogue response generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31.

Sharma, D.; and Bikshandi, B. 2020. Artificial Empathy–An Artificial Intelligence Challenge. In *Artificial Intelligence*, 321–326. Productivity Press.

Wu, Q.; Li, L.; and Yu, Z. 2020. TextGAIL: Generative adversarial imitation learning for text generation. *arXiv preprint arXiv:2004.13796*.

Zhang, Y.; Wang, Y.; Zhang, L.; Zhang, Z.; and Gai, K. 2019. Improve diverse text generation by self labeling conditional variational auto encoder. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2767–2771. IEEE.

Appendix

Algorithm 1 Modified GAIL Algorithm (EmpathyGAIL)

Input: Expert empathetic trajectories τ_E each with history h_E, prompt p_E and response r_E; initial policy and discriminator parameters θ_0, w_0

for $i = 0, 1, 2, \ldots$ **do**
  Sample trajectories τ_i for discriminator D
  Collect occupancy score ρ_i from discriminator D for all τ_i
  Update discriminator parameters from w_i to w_{i+1}
  Update generator parameters by taking policy step from θ_i to θ_{i+1} using PPO
end for

Algorithm 2 Generator Algorithm within modified GAIL algorithm

Input: Initial policy parameters θ_0, Noise prior $p(z)$

for $i = 0, 1, 2, \ldots$ **do**
  Sample noise sample $z_i \sim p(z)$
  Generate possible responses using GPT-2 language model
  Recover generated response using z_i to form γ_i
  Update parameters by taking policy step from θ_i to θ_{i+1} using PPO
end for

Output: Generated policy γ_G each with history h_G, prompt p_G, and response r_G

Hyperparameter	**Value**
Sample Batch Size	16
Generator Warmup Scheduler	WarmupLinear
Generator Warmup Steps	1000
Human Demo Ratio	0.3
Human Demo Ratio Warmup Steps	100
PPO Buffer Size	128
PPO Minibatch Size	8
PPO Epsilon	0.2
Discriminator Pre-Train Steps	200
Discriminator Optimizer	AdamW
Discriminator Optimizer Learning Rate	1×10^{-4}
Weight Decay	0.01
Layer Normalization Epsilon	1×10^{-5}
Activation Function	tanh
Epochs	750
Batch Size	32
Validation Frequency	10

Table 3: Final EmpathyGAIL model hyperparameter values of the model using expert empathetic trajectories in the form (h_i, p_i, r_i) after tuning.

Metric	Model	1 Turn	2 Turn	3+ Turn
Perplexity	GPT-2 + MLE	32.54 ± 2.97	24.14 ± 2.19	18.11 ± 1.89
	RL-Seq2Seq	97.19 ± 14.10	71.43 ± 10.10	60.29 ± 9.20
	TextGAIL	10.34 ± 2.11	7.13 ± 1.74	5.18 ± 1.29
	EmpathyGAIL	$\mathbf{6.24 \pm 1.14}$	$\mathbf{4.31 \pm 0.95}$	$\mathbf{3.97 \pm 0.73}$
BLEU	GPT-2 + MLE	8.24 ± 0.15	15.39 ± 0.34	22.14 ± 0.23
	RL-Seq2Seq	3.23 ± 0.03	4.51 ± 0.05	4.41 ± 0.04
	TextGAIL	22.01 ± 0.24	27.34 ± 0.33	30.59 ± 0.44
	EmpathyGAIL	$\mathbf{28.91 \pm 0.44}$	$\mathbf{33.50 \pm 0.25}$	$\mathbf{39.20 \pm 0.81}$

Table 4: Perplexity and BLEU Error Results trained on only EmpatheticDialogues expert trajectories for 1 Turn, 2 Turn, and 3+ Turn Conversations

Metric	Model	1 Turn	2 Turn	3+ Turn
Perplexity	GPT-2 + MLE	38.24 ± 3.61	33.14 ± 4.13	16.13 ± 2.93
	RL-Seq2Seq	124.16 ± 14.69	84.03 ± 13.23	59.03 ± 9.94
	TextGAIL	15.35 ± 3.81	7.49 ± 2.35	5.14 ± 1.35
	EmpathyGAIL	$\mathbf{7.29 \pm 1.49}$	$\mathbf{5.00 \pm 0.99}$	$\mathbf{3.41 \pm 0.48}$
BLEU	GPT-2 + MLE	6.23 ± 0.05	8.10 ± 0.33	13.22 ± 0.98
	RL-Seq2Seq	1.87 ± 0.09	2.02 ± 0.07	3.99 ± 0.14
	TextGAIL	10.34 ± 0.44	15.91 ± 0.67	19.94 ± 0.24
	EmpathyGAIL	$\mathbf{14.14 \pm 0.31}$	$\mathbf{20.11 \pm 0.39}$	$\mathbf{27.40 \pm 0.59}$

Table 5: Perplexity and BLEU Error Results trained on only DailyDialog expert trajectories for 1 Turn, 2 Turn, and 3+ Turn Conversations

Metric	Model	1 Turn	2 Turn	3+ Turn
Perplexity	GPT-2 + MLE	59.34 ± 5.19	48.22 ± 4.10	34.11 ± 3.47
	RL-Seq2Seq	158.01 ± 23.94	113.10 ± 18.18	103.20 ± 14.81
	TextGAIL	23.10 ± 4.64	19.01 ± 3.12	10.33 ± 3.05
	EmpathyGAIL	$\mathbf{18.34 \pm 4.19}$	$\mathbf{12.13 \pm 3.59}$	$\mathbf{8.18 \pm 2.81}$
BLEU	GPT-2 + MLE	2.01 ± 0.03	3.59 ± 0.07	5.54 ± 0.14
	RL-Seq2Seq	0.34 ± 0.01	0.62 ± 0.03	0.74 ± 0.02
	TextGAIL	6.96 ± 0.49	10.31 ± 0.50	14.27 ± 0.78
	EmpathyGAIL	$\mathbf{8.24 \pm 0.59}$	$\mathbf{14.46 \pm 0.93}$	$\mathbf{20.21 \pm 0.88}$

Table 6: Perplexity and BLEU Error Results trained on only empathetic tweet expert trajectories for 1 Turn, 2 Turn, and 3+ Turn Conversations

	Example Results
History	I moved to LA a few years ago and I never had the chance to visit home.
Prompt	I'm going to see my parents soon!
EmpathyGAIL	That's awesome! I hope you have a good time.
TextGAIL	Oh, it seems you're getting pretty into it.
RL-Seq2Seq	I only am here due to my great parents who are in Australia.
GPT-2 + MLE	I'm going to see my mom soon.
Human Response	Great! I'm glad you're happy.
History	None
Prompt	He was a very good friend of mine. Last month, he passed away.
EmpathyGAIL	I am very sad to hear that he is no longe
TextGAIL	I'm sorry for your loss.
RL-Seq2Seq	I used to be a good friend to many of our fans and the right of someone.
GPT-2 + MLE	Just wanted to say thank you from the bottom of my heart. I will miss him very much.
Human Response	I hope you are doing OK.
History	I was in a drive-thru line and someone cut in front of me.
Prompt	I was furious and blew up at him.
EmpathyGAIL	Sometimes it makes more sense to just walk away. You cannot control others, you can only control yourself.
TextGAIL	I was furious and laughed at this.
RL-Seq2Seq	When I had had so much fun, we went off together, it took me, in spite of what.
GPT-2 + MLE	I was furious and blew up at him.
Human Response	I find that meditating helps calm me down.
History	None
Prompt	I had a job interview today and I think it went really well.
EmpathyGAIL	Wonderful!
TextGAIL	I was really happy and really, really glad.
RL-Seq2Seq	I can now say that I am now a professional, something I couldn't say even a year ago.
GPT-2 + MLE	I'm so glad I got to work today.
Human Response	That's nice to hear!

Table 7: Extended Qualitative Model Results against Baselines (TextGAIL, RL-Seq2Seq, GPT-2 + MLE)