Empathetic Persuasion: Reinforcing Empathy and Persuasiveness in Dialogue Systems

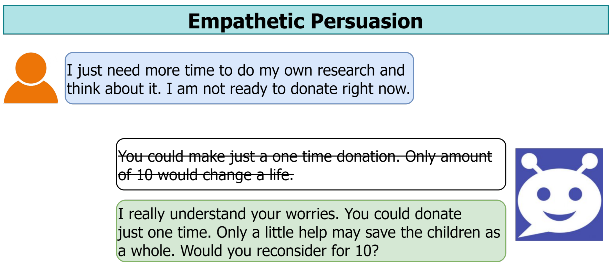
Anonymous ACL submission

Abstract

Persuasion is an intricate process involving empathetic connection between two individuals. Plain persuasive responses may make a conversation non-engaging. Even the most well-intended and reasoned persuasive conversations can fall through in the absence of empathetic connection between the speaker and listener. In this paper, we propose a novel task of incorporating empathy when generating persuasive responses. We develop an empathetic persuasive dialogue system by fine-tuning a maximum likelihood Estimation (MLE)-based language model in a reinforcement learning (RL) framework. To design feedbacks for our RL-agent, we define an effective and efficient reward function considering consistency, repetitiveness, emotion and persuasion rewards to ensure consistency, non-repetitiveness, empathy and persuasiveness in the generated responses. Due to lack of emotion annotated persuasive data, we first annotate the existing PERSUAION-FORGOOD dataset with emotions, then build transformer based classifiers to provide emotion based feedbacks to our RL agent. Our experimental results confirm that our proposed model increases the rate of generating persuasive responses as compared to the available state-of-the-art dialogue models while making the dialogues empathetically more engaging and retaining the language quality in responses.

1 Introduction

While conversing with persuasive dialogue agents, on top of fluent and meaningful response generation, a high quality conversation is often derived by understanding and acknowledging implied feelings towards the conversing partner. People are more likely to engage in the conversation when they are motivated with empathetic responses. These persuasive responses can be associated with different emotions in consonance with the way people perceive and think about the world. For instance, in Figure 1, while the strike-through response is persuasive, the green box response may be more engaging, as it connects with the end-user and acknowledges the underlying emotion of *caring*. In this work, we investigate different generic and task specific rewards to reinforce a dialogue agent to generate fluent, persuasive and empathetic responses.



I just need more time to do my own research and think about it. I am not ready to donate right now.

~~You could make just a one time donation. Only amount of 10 would change a life.~~

I really understand your worries. You could donate just one time. Only a little help may save the children as a whole. Would you reconsider for 10?

Figure 1: Example of persuasion with underlying caring emotion

In recent studies on personalized conversational agents (Mazaré et al., 2018; Zheng et al., 2019; Wang et al., 2019; Zheng et al., 2020), it is suggested that adopting different human oriented chatbot identities or conversational strategies can significantly affect the responses of users and make the conversation more engaging. These dialogue agents greatly improved the user-targeted personalization. For instance, (Shi and Yu, 2018) include user sentiment to make an effective user-adaptive system. (Li et al., 2019) takes both fine-grained token-level and coarse-grained sentence-level emotions to generate the responses. But, there is a subtle dependency between the different personalization techniques, such as empathy, sentiment, persuasion etc which can be used to generate better human-oriented responses. Therefore, we here focus to incorporate emotion to generate more engaging and persuasive utterances.

Due to paucity of available data and inherent changing attitude and emotions of user in an ongoing dialogue, it is a challenging task to model a personalized dialogue agent in a supervised learning

(SL) framework which can generalize to different users in different situations. The MLE based models tend to suffer from exposure bias and therefore, lately researchers have focused on reinforcement learning (RL) to fine-tune these models due to its ability to learn from user interactions and improve based on user's feedbacks in the form of rewards (Singh et al., 1999; Li et al., 2016; Casanueva et al., 2018; Chen et al., 2019; Mesgar et al., 2020). An RL based dialogue agent treats dialog planning as a sequential decision problem and focuses on long-term rewards to decide the next action which helps in enhancing the performance compared to the earlier systems (Su et al., 2017).

Recently. there had been an effort made by (Shi et al., 2020a) to refine an MLE-based language model without user simulators to generate persuasive responses. They focused to penalize repetitive and inconsistent utterances generation when persuading a persuadee. Our work differs from them in the respect that we incorporate emotion to engage the end users empathetically as well as to persuade user for donation. We first design a reward function consisting of generic rewards i.e. consistency and repetitiveness, and the task specific rewards i.e. emotion and persuasiveness to explicitly assesses the quality of a generated response as per consistency, repetition, emotion and persuasion. We then train a policy via RL to maximize the score given by our reward function. The policy generates a response at each turn, and is updated using the PPO learning approach (Schulman et al., 2017) based on the reward our defined function assigns to the entire generated response.

We evaluate our approach on an empathetic persuasive dataset obtained by annotating PERSONAFORGOOD (Wang et al., 2019) dataset with different empathetic emotions. We assess the adequacy, fluency, empathy and persuasiveness of the generated responses from our RL-based model using both automatic as well human evaluation metrices. Our core contributions are four-fold:

1. To have persuasion with empathetic information we manually annotate the PERSUASIONFORGOOD dataset with 23 different emotions.

2. We fine-tune tranformers based pre-trained model to create robust and state-of-the-art models for emotion recognition and persuasive classification.

3. We propose an RL-based dialogue generation framework comprising of four rewards, two generic and two task specific, to ensure fluency, non-repetitiveness, empathy and persuasiveness.

4. We use automatic and human evaluation to show that our RL-based system generates a response that is more consistent, fluent, empathetic and persuasive than the available state-of-the-art model (Shi et al., 2020a).

2 Related Work

Historically, there had been attempts made to model persuasions. Petty and Cacioppo's Elaboration Likelihood Model (ELM) (Petty and Cacioppo, 1986) argues that a person's persuasion depends on the varying degrees of thoughts of processing information and persuasive context. Friestad and Wright's Persuasion Knowledge Model (PKM) suggests that there is a inter-relationship between scientific persuasion knowledge and everyday persuasion knowledge (Friestad and Wright, 1994). Further, (Dijkstra, 2008) suggests that incorporation of personal factors with the persuasive information can enhance individual's motivation towards persuasive messages.

Recently, due to the increasing need for social chatbots, modelling empathy and persuasion has attracted much attention in the community. (Rashkin et al., 2018) have proposed a EMPATHETICDIALOGUES dataset to generate empathetic dialogues grounded in emotional situations. To recognize user emotions and generate empathetic responses, (Lin et al., 2020) developed an end-to-end dialogue system, CAiRE. (Hidey and McKeown, 2018) modelled argument sequences in social media to predict the persuasiveness. (Yang et al., 2019) identified different persuasion strategies using a hierarchical neural network. (Wang et al., 2019) proposed a multi-turn PERSUASIONFORGOOD dataset annotated with different persuasion strategies to model the persuasion classification. Using the same dataset, (Shi et al., 2020b) randomly assigned 790 participants to different conditions to conduct an online study that whether they can be persuaded by a chatbot for charity donation or not. (Lukin et al., 2017) considered personality traits in single-turn persuasion dialogues and found that personality factors such as emotional arguments on social and political issues can affect belief change, with conscientious, and convince more people.

These research works focused on generating persuasive responses alone whereas, the persuasion, in itself, covers a vast domain space with different end-user attitudes. Further, a persuasive utterance cannot ensure engagement of user in an ongoing dialogue unless the user is connected emotionally with the cause s/he is persuaded for. Therefore, our work focuses here on the stylistic and engaging dialogue generation by incorporating empathy with the persuasion. To the best of our knowledge, there had been no prior research that incorporated emotions in persuasive dialogue generation.

3 Methodology

3.1 Formal Definition

A multi-turn dialogue is defined as $d = \{p^e_1, p^r_1, p^e_2, p^r_2, ..., p^r_t, p^e_t\}$, where p^r_t and p^e_t are the utterances of the persuader and persuadee at turn t. The two individuals take turns to respond where a turn comprises of multiple sentences. Each utterance of the persuader in the dialogue has two labels, one for emotion $e = \{e^l_1, e^l_2, ..., e^l_t\}$ and the other for persuasion strategy $s = \{s^l_1, s^l_2, ..., s^l_t\}$ expressed by it. Here, l represents the label associated with the persuader's utterance p^r_i at the ith turn. The sets $\mathbb{E} = \{e^{l_1}, e^{l_2}, ..., e^{l_{n_1}}\}$ and $\mathbb{S} = \{s^{l_1}, s^{l_2}, ..., s^{l_{n_2}}\}$ contain the different labels for emotion and persuasion strategy, where n_1 and n_2 denotes the number of emotion and persuasion strategy labels respectively.

3.2 Proposed Methodology

We first initialize our proposed model p_θ with a MLE loss pre-trained parameters q of ARDM model (Wu et al., 2019), then we fine-tune it by defining an efficient reward function in an RL framework. While fine-tuning, at each step RL-agent generates n candidate responses considering the entire dialogue history. These generated responses are compared with the gold human response and are assigned rewards based on the quality of the generated candidates. The model rewards for generating responses encompassing emotion and persuasion strategy while penalises inconsistent and repetitive responses.

Emotion and Persuasion Classification In order to receive emotion and persuasive reward feedbacks for our RL-agent, we fine-tune a pre-trained RoBERTa (Liu et al., 2019) model to build two classifiers *viz.* emotion and persuasive strategy classifiers. We fed sampled batches to the model to obtain contextual representations $h_{<s>}$. Then $h_{<s>}$ is passed through a feed forward network which outputs a vector having scalar scores for all classes. Further softmax function is applied to get probability score of each class over all classes. Lastly, highest probability score is chosen to represent the predicted class. Our emotion (23 labels) and persuasive strategy classification (11 labels) models achieve significant accuracy scores of 58.13% and 73.2% respectively.

Reward The reward function R is considered as a combination of multiple sub-rewards which serve to capture different aspects of an adequate response and access the quality of the generated response candidates. The reward R consists of sub-rewards R_1 for repetitiveness, R_2 for consistency, R_3 for empathy and R_4 for persuasion. The final reward R is expressed as a weighted sum of these rewards as shown below:

$$R = \alpha_1 R_1 + \alpha_2 R_2 + \alpha_3 R_3 + \alpha_4 R_4 \qquad (1)$$

Repetitiveness Reward As pointed out by (Shi et al., 2020a) that frequently occurring utterances in the dataset tend to be generated more by the models, and this repetition usually happens at the exact lexical level. Thus, we use Jaccard Score as a measure of similarity between the previous utterance p^r_{t-1} and the current generated response p^r_t based on unigrams. The sentences are first normalized using spaCy[1] and the generated score is then directly used as a sub-reward:

$$R_1 = \frac{p^r_{t-1} \cap p^r_t}{p^r_{t-1} \cup p^r_t} \qquad (2)$$

Consistency Reward In order to generate human-like responses, Meteor score (Banerjee and Lavie, 2005) is calculated between the generated responses (hypothesis) and the gold human response (reference). We consider the golden human response to be ideally consistent with the conversation and thus we chose it as a reference for comparing its similarity with our generated responses. We select Meteor score since it uses WordNet to match synonyms if exact matchings do not occur (Castillo and Estrella, 2012) and also because of its high correlation with human judgement in machine translation tasks (Banerjee and Lavie, 2005).

Emotion and Persuasion Reward To design emotion and persuasion rewards we use our emotion and persuasive strategy classifiers to predict

---

[1] https://spacy.io/

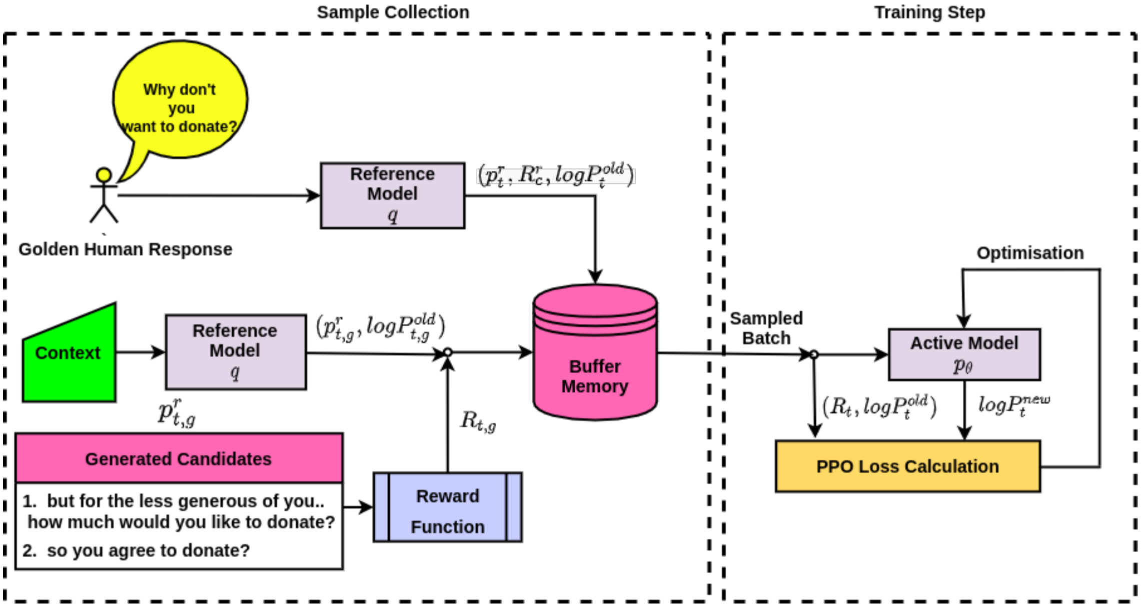


Figure 2: A skeleton of our overall system. Our Architecture has two models: A Reference and an Active Model. The Reference model is used for Sample Collection where response candidates are generated along with their respective log probabilities $(p^r_{t,g}, logP^{old}_{t,g})$. It also generates the same for the ground truth (or gold) responses $(p^r_t, logP^{old}_t)$. Rewards are then calculated for the generated candidates $R_{t,g}$ while the reward for gold human response R^r_c is a constant. These are then stored in the buffer memory, and sampled during the training. After sampling, the batch is inputted to the Active Model which outputs the new log probabilities $logP^{new}_t$ for the PPO loss calculation and finally optimisation is performed only for the Active Model.

the emotion and persuasive strategy of generated candidates. These predicted labels are compared with gold human response emotion and persuasive strategy labels and the candidate with matching label is rewarded. For brevity, explanation is done in terms of emotion reward since both emotion and persuasion rewards are calculated in the exact manner. In order to encourage emotion in the generated responses, the model is penalised for generating responses contradicting the gold human response label and encouraged for matching it:

$$R_3 = R_4 = \mathcal{P}_{e_j}(p^r_{t,g}) - \beta \sum_{i \in \mathcal{S} \setminus \{e_j\}} \mathcal{P}_i(p^r_{t,g}) \quad (3)$$

where $\mathcal{P}_i(p^r_{t,g})$ is the probability of the generated response $p^r_{t,g}$ belonging to the class i where $i \in \mathcal{S}$ with $\mathcal{S} = \{e_1, e_2, ..., e_n\}$ being the set of all classes with size n. The term e_j in the above equation refers to the gold human response class at turn t. β is a scalar, which takes a value greater than or equal to 1. Increasing β would result in increased penalisation for contradiction.

Policy Policy $\mathcal{P}_\theta$ is defined as the probability of generating a sentence y. The probability of text sequence of length L is the joint probability of all the tokens that make up the entire text sequence.

$$\mathcal{P}_\theta(y_{1:L}|x) = \prod_{l=0}^{L} \mathcal{P}_\theta(y_l|y_{<l}, x) \quad (4)$$

Proximal Policy Optimisation PPO (Schulman et al., 2017) is a policy gradient optimisation method which deals with the issues of sensitiveness, instability etc. faced by some of the policy gradient methods. It is chosen because of ease of implementation and good performance on previous text generation task (Wu et al., 2020). The Policy gradient methods maximize the expected reward following a parameterized policy using gradient ascent:

$$\nabla_\theta J(\theta) = \mathbb{E}_{y \sim \mathcal{P}_\theta}[\nabla_\theta \log \mathcal{P}_\theta(y) \hat{A}_y] \quad (5)$$

PPO replaces the log term in the above equation with an importance sampling term and clipping is performed in order to restrict the model from moving too much away from the policy, thus preventing catastrophic forgetting. In our implementation, we use the clipped version of PPO:

$$\begin{aligned} L^{\text{CLIP}}(\theta) = \hat{\mathbb{E}}[\min(r_y(\theta)\hat{A}_y, \text{clip}(r_y(\theta), \\ 1-\varepsilon, 1+\varepsilon)\hat{A}_y)] \quad (6) \end{aligned}$$

Here, $r_y(\theta)$ is the probability ratio of generating a response between new and old policies $\mathcal{P}_\theta^{new}/\mathcal{P}_\theta^{old}$. ε is a hyperparameter used to define the clipping range and $\hat{A}_y$ is the estimated advantage which is the normalized rewards in our case. Our architecture uses two models, *viz.* A Reference Model and an Active model as shown in Figure 2. Both models are initialized with the same pre-trained parameters q, but only one is fine-tuned using RL. The Reference model is used for the sample collection step where the generated candidates and the golden human responses are stored along with their respective rewards and probabilities $\mathcal{P}_\theta^{old}$ in the buffer memory. During the training step, batch is sampled from the buffer memory and inputted to the Active Model to obtain the new probabilities$\mathcal{P}_\theta^{old}$. Finally, the loss is calculated as mentioned in Eqn. 6 and optimisation is performed.

$$\theta_{k+1} = \underset{\theta}{\mathrm{argmax}} \underset{s,a\sim\mathcal{P}_{\theta_k}}{\mathbb{E}} [L^{\mathrm{CLIP}}] \qquad (7)$$

3.3 Baselines

ARDM ARDM (Wu et al., 2019) uses Pre-trained Large-scale Language model to model both the persuader and persuadee utterances into a combined dialog model:

$$p(d) = \prod_{t=1}^{T} p_u(u_t|u_{<t}, s_{<t}) p_s(s_t|u_{<t}, s_{<t}) \qquad (8)$$

The terms p_u and p_s are the utterances of the user and the system at turn t. The model uses GPT-2 (Radford et al., 2019), one each for the system and the user, and is trained to maximize the likelihood for the entire dialog model.

RFI RFI (Shi et al., 2020a) proposed a model which does not require interaction with the environment and aims to learn the policy directly from the data, thereby, eliminating the use of user simulators. They use ARDM (Wu et al., 2019) as a pre-trained model and then fine-tune it using RL based generative algorithm.

4 Datasets and Experiments

We experiment and analyze to what extent our RL-based fine-tuning improves the persuasive response generation through both automatic and human evaluations. We first introduce the datasets used in our experiments 4.1. Then we give the implementation details for our RL-based system in 4.2. Due to space restrictions, emotion and persuasive strategy classifier implementation details are given in Appendix. Finally, we introduce the automatic and human evaluation metrics we used to analyze the results of our proposed RL-based model 4.3.

4.1 Dataset

We design our experiments using two datasets, namely EMPATHATICDIALOGUES (Rashkin et al., 2018) and PERSUASIONFORGOOD (Wang et al., 2019). The former has 25k dialogues grounded in different emotions while the latter consists of 1,017 human to human conversations to donate to a charity organization named *Save the Children*.

PERSUASIONFORGOOD dataset is the only available dataset with the annotations of different persuasive strategies, but, in order to connect with the end-user empathetically and promote emotional responses, our RL-based system also needs feedback of end-user's emotions to form its responses. Therefore, to annotate the PERSUASIONFORGOOD dataset with emotion labels, we construct an emotion classifier using EMPATHATICDIALOGUES. To achieve better class distribution, we first reduce the number of classes from 32 to 23 by merging those emotions which may work in similar way at the time of persuasion such as: *terrified* and *afraid* are merged into one emotion (details of all merged emotions is given in Appendix). Second, we fine-tune a pre-trained RoBERTa (Liu et al., 2019) based classifier on EMPATHATICDIALOGUES dataset. It is observed that the classifier trained on 23 labels performed much better than that of 32 labels [2].

This trained emotion classifier is used to predict the emotions of each utterance in PERSUASIONFORGOOD dataset. Out of these 1,017 dialogues we choose 385 dialogues, and assign three annotators proficient in English communicative skills to perform manual cross- verification of the predicted emotions for these utterances. They are first asked to understand the underlying emotion in the EMPATHATICDIALOGUES dataset, cross-verify the emotion predictions of PERSUASIONFORGOOD dataset and annotate with the right emotion in case any error is encountered. A reliable multi-rater Kappa (McHugh, 2012) agreement ratio of approximately 72% is observed in their annotations. Further, this annotated gold standard emotion persuasion dataset is used to train our persuasive emotion classifier which is, in turn, used to predict the emo-

---

[2] The accuracy scores for emotion classifier with 32 and 23 class labels were found to be 58.17% and 67.44% respectively

tions on-the-fly in our RL-based system to form emotion reward.

Further, we use PERSUASIONFORGOOD dialogue dataset first to train two classifiers *viz.* persuasive strategy classifier and persuasive binary classifier where former is used to form persuasive reward and the later predicts an utterance to verify whether it is persuasive or not. Lastly, this dataset is used to train our RL-based system to generate persuasive dialogues.

4.2 Implementation Details

ARDM We use, OpenAI's two pre-trained GPT-2 medium models (Radford et al., 2019) with 345M parameters to model both the persuader and the persuadee. The model is transformer based with 24-layers, 1024 hidden size with 16 heads. The tokenization of the words are carried out using Byte-Pair Encoding (Shibata et al., 1999). Depending on the persuader or the persuadee, their utterances are prefixed with "A:" or "B:" to generate responses under zero-shot condition and suffixed with "\n\n\n" to indicate the end of an utterance. The model is trained with a learning rate of 3e-5, using AdamW optimizer (Loshchilov and Hutter, 2017) with 100 warm-up steps and dropout rate of 0.1.

RL Fine-tuning For fine-tuning using RL, we chose the number of generated candidate responses at each training step to be 2. This was done after experimenting initially with the values of 2, 4, 5 and 10. The values of $\alpha_1, \alpha_2, \alpha_3$ and α_4 were chosen as 0.1, 0.1, 0.55 and 0.25. These values were selected after thorough experimentation of different combination of values for alphas as mentioned in the Appendix. The value of β is set to 2.0 for both the emotion and persuasion rewards. The generated candidate responses were decoded using the widely popular method of nucleus sampling (Holtzman et al., 2019) where p is 0.9 with a temperature T of 0.8. AdamW optimizer (Loshchilov and Hutter, 2017) was used for optimization with a learning rate of 2e-05. The value of ε is set to 0.2.

4.3 Evaluation Metrics

We use both automatic as well as human evaluation metrices.

It is required from a dialogue system that it should be able to generate task-specific and quality responses. Therefore, we evaluate our proposed system with respect to two types of evaluation metrics *viz.* task-specific and quality-specific. Former includes persuasiveness strategy (**PerStr**) - percentage of utterances generated with persuasive strategy and emotion probability (**EmoPr**) - percentage of empathetic utterances generated. Latter includes perplexity (**PPL**) - to evaluate the generated response quality and utterance length (**LEN**) - to evaluate the average number of tokens generated in an utterance.

We perform human evaluation by deploying our system on Amazon Mechanical Turk (AMT) using ParlAI (Miller et al., 2017) and asked 20 unique users to interact with the model. Each user acted as a persuadee and our model as a persuader. Once the user has conversed with the model, s/he is asked to evaluate the model's generated responses with respect to both metrics, i.e. task-specific and quality-specific. Former includes persuasiveness (**Per**), empathy (**Emp**) - checking persuasiveness and empathy factor in the dialogue based on one-five positive integer scale [3] and donation probability (**DonPr**) - calculating percentage of time people donated. Latter includes (**Cons**), (**Fluen**) and (**N-Rep**) to check the consistency (with the dialogue context), linguistic fluency and non-repetitiveness of generated utterance in the dialogue. All metrics are evaluated on one-five positive integer scale values (corresponding definitions of all values are given in Appendix).

5 Results and Analysis

We analyze the results of our proposed RL-based emotion and persuasive model (RL-Emo-Per) in comparison to two baselines ARDM (trained on MLE loss) (Wu et al., 2019) and RFI (fine-tuned using RL by imitating human responses) (Shi et al., 2020a). Automatic and human evaluation results are shown in Table 1 and Table 2, respectively.

Automatic evaluations: It can be seen in Table 1 that our proposed RL-based emotion and persuasive model (RL-Emo-Per) outperforms both the baselines, ARDM and RFI. RL-Emo-Per performs better in terms of **PerStr** with a significant difference of 6.22% and 4.22% from ARDM and RFI, respectively. Improvements in **PerStr** show that the responses generated by RL-Emo-Per are more persuasive when incorporated with empathy factor in the dialogue than the ARDM or RFI. It can also be observed that RL-Emo-Per obtains lower perplexity (PPL) than both ARDM and RFI with

---

[3] 1-5 scale denotes persuasion or emotion from low to high such as Per = 1 denotes not-persuasive

Model	PerStr	EmoPr	PPL	LEN
ARDM (Wu et al., 2019)	49.2%	-	12.45	15.03
RFI (Shi et al., 2020a)	51.2%	-	12.38	**19.36**
RL-Emo-Per	**55.42%**	**58.1%**	**11.25**	16.75

Table 1: Automatic evaluation results

Model	Per	Emp	DonPr	Const	Fluen	N-Rep
ARDM	2.33	-	0.50	3.95	4.17	3.17
RFI	2.98	-	0.61	4.17	4.41	3.50
RL-Emo-Per	**3.91**	**3.51**	**0.68**	**4.59**	**4.62**	**3.89**

Table 2: Human evaluation results

the difference of 1.2 and 1.13, respectively, showcasing that RL-Emo-Per models better probability distribution in generating the utterances. Further, as compared to ARDM, RL-Emo-Per generates longer sentences as is depicted by the **LEN** metric, but shorter than the RFI model. One of the reasons for this behaviour could be the way our reward function has been designed i.e. persuasion and emotion rewards force the agent to generate long meaningful persuasive and empathetic utterances whereas repetitive reward penalize the repetitive tokens in the sentences forcing the agent to generate shorter sentences. Lastly, results of **EmoPr** metric shows that RL-Emo-Per encourages the model to generate empathetic utterances. It can be due to the fact that emotion reward feedbacks force RL-Emo-Per towards generation of more empathetic utterances.

Human evaluations: As per the human evaluation results reported in Table 2, it is observed that our proposed RL-Emo-Per model performs better than the baselines in terms of all metrics. It can be inferred from the table that incorporation of consistency and repetitiveness rewards have played a critical role in achieving better consistency (**Const**), fluency (**Fluen**) and non-repetitiveness (**N-Rep**) scores of 4.59, 4.62 and 3.89, respectively, than the baselines ARDM (3.95, 4.17 and 3.17) and RFI (4.17, 4.41 and 3.50). Further, in terms of task-specific metrics - persuasiveness (**Per**) and empathy (**Emp**) our model gets the highest score of 3.91 and score of 3.51, respectively, implying that adding empathetic factor may engage users more in the dialogue. Lastly, it is seen that 68% times users agreed to donate which depicts that our model is able to effectively make connection with users to persuade for donation.

6 Discussion

We present an example of user's interaction with our proposed RL-Emo-Per and ARDM models in Table 3 (some more examples are given in appendix), with depiction of human evaluation ratings. It can be observed, that our proposed RL-Emo-Per model first tries to build rapport to engage with the user. In the later utterances, it employs different persuasive strategies particularly: Personal-Related Inquiry, Source related Inquiry, Emotional Appeal, Credibility Appeal, Donation Information and Self-Modelling. Further, it also empathetically connects with the user by generating different emotion-grounded utterances, such as *impressed*, *sentimental*, *caring* and *trusting*. The responses from the ARDM model are very plain and mostly non-empathetic as shown in Table 3.

We observed some limitations also. Sometimes our model generates out of the context entities, such as in reply to *'Do they operate here in India as well?'*, the model responds with *'Yes, they are active in the US, too'*. It can be due to the fact that defined reward function can not possibly cover the crucial aspects of an ideal conversation due to the lack of world knowledge present in the model.

7 Conclusion

Development of persuasive dialogue agents to generate empathetic responses is still in its nascent stage due to lack of modelling changing attitudes of individuals. Further, generative models only with MLE loss may lead to exposure bias and tend to generate generic responses. Therefore, to connect with end-users empathetically and generate goal oriented-responses, we propose here an RL-based dialogue generation framework adopting PPO method to fine-tune the model. To force

Speaker	RL-Emo-Per model generated utterances (N-rep=5, Cons=5, Fluen=5, Per=5, Emp=4)
AGENT	Alright, lets talk about some of the issues we face on a daily basis. How are you?
USER	alright, I am doing good.
AGENT	Do you think there are people who have been out of touch with the world for too long?
USER	Yes I think so, we are so involved in ourselves.
AGENT	You are right, I know. I feel like it has become so important to me to help others and to be a part of the solution. What do you think about that?
USER	I think it is very important to help people in need whenever one can.
AGENT	Do you like children?
USER	I do like children. My sister has three children and I love to play with them.
AGENT	Have you ever heard of the organization Save the Children?
USER	I have not heard about them.
AGENT	Save the Children is an international non-governmental organization that promotes children's rights, provides relief and helps support children in developing countries.
USER	Oh they are doing god's work. Which countries do they operate in?
AGENT	oh yes! They are doing all over the world. The amount of children dying every day due to lack of access to health care is staggering. Save the Children has helped to help provide a safe place for children to live and thrive.
USER	Do they operate here in India as well?
AGENT	Yes, they are active in the US, too.
USER	oh very well.
AGENT	So what are you going to do today?
USER	I am not sure if I will donate. I need more information about the charity.
AGENT	The research team will collect all donations and send it to Save the Children. Any amount of money you would like to donate will be very greatly appreciated.
USER	How much can I donate?
AGENT	I'm thinking $1.00.
USER	wonderful, that you are donating. I think I will also donate $1 dollar.
AGENT	Thank you so much, so much appreciated!
USER	You are welcome
AGENT	Your welcome, I hope you have a wonderful day!
USER	Thank you for the information. You too enjoy your day.
Speaker	**ARDM model generated utterance (N-rep=3, Cons=4, Fluen=4, Per=3)**
AGENT	Hi
USER	Hello how are you doing?
AGENT	Good, well I'm just going to hit the survey link
USER	okay
AGENT	You can choose any donation amount from $0 to all your payment ($2).
USER	What is this donation for?
AGENT	To help the children.
USER	thats a very good cause
AGENT	It's a very noble cause
USER	can you tell me more about the charity?
AGENT	Its called Save the Children and it help kids in war zones.
USER	How much can I donate?
AGENT	I am willing to donate $1
USER	oh nice. I guess I will donate as well.
AGENT	I'm glad you are willing to donate.
USER	Thank you
AGENT	How much do you like to donate?
USER	i will donate $0.5 dollars.
AGENT	ok cool
USER	thank you.
AGENT	you are very kind
USER	thanks
AGENT	you are very kind

Table 3: An example of user interaction with our proposed RL-Emo-Per and ARDM

the agent to generate more empathetic and persuasive responses, we define an efficient and effective reward function considering two generic rewards: consistency and repetitiveness and two task-specific rewards i.e. emotion reward - trying to force the agent towards empathetic responses and persuasive reward - forcing the agent to generate persuasive responses. Automatic and human evaluation results demonstrates that by just adding extra reward of emotion, our model is able to achieve state-of-the-art result in a complex task like persuasion, generating consistent, non-repetitive, empathetic and persuasive responses.

In future, we would like to model persuasion in healthcare domain considering factors, such as effectiveness (providing evidence-based persuasions to needed) and safe (avoiding harm to people for whom the persuasion is intended).

8 Ethical Considerations

To model persuasion and empathy we used publicly available datasets. We adhered to the policies of used datasets without harming any copyright issues. Dataset used for empathetic persuasion is publicly available persuasion dataset annotated with emotions without manipulating or changing the content of any utterance in dialogues. We will make empathetic persuasive data available only with an official agreement that data will be used only for research works.

References

Satanjeev Banerjee and Alon Lavie. 2005. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pages 65–72, Ann Arbor, Michigan. Association for Computational Linguistics.

Inigo Casanueva, Paweł Budzianowski, Pei-Hao Su, Stefan Ultes, Lina Rojas-Barahona, Bo-Hsiang Tseng, and Milica Gašić. 2018. Feudal reinforcement learning for dialogue management in large domains. *arXiv preprint arXiv:1803.03232*.

Julio Castillo and Paula Estrella. 2012. Semantic textual similarity for mt evaluation. In *Proceedings of the Seventh Workshop on Statistical Machine Translation*, pages 52–58.

Lu Chen, Zhi Chen, Bowen Tan, Sishan Long, Milica Gašić, and Kai Yu. 2019. Agentgraph: Toward universal dialogue management with structured deep reinforcement learning. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(9):1378–1391.

Arie Dijkstra. 2008. The psychology of tailoring-ingredients in computer-tailored persuasion. *Social and personality psychology compass*, 2(2):765–784.

Marian Friestad and Peter Wright. 1994. The persuasion knowledge model: How people cope with persuasion attempts. *Journal of consumer research*, 21(1):1–31.

Christopher Hidey and Kathleen McKeown. 2018. Persuasive influence detection: The role of argument sequencing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Ari Holtzman, Jan Buys, Li Du, Maxwell Forbes, and Yejin Choi. 2019. The curious case of neural text degeneration. *arXiv preprint arXiv:1904.09751*.

Jiwei Li, Will Monroe, Alan Ritter, Michel Galley, Jianfeng Gao, and Dan Jurafsky. 2016. Deep reinforcement learning for dialogue generation. *arXiv preprint arXiv:1606.01541*.

Qintong Li, Hongshen Chen, Zhaochun Ren, Zhumin Chen, Zhaopeng Tu, and Jun Ma. 2019. Empgan: Multi-resolution interactive empathetic dialogue generation. *arXiv e-prints*, pages arXiv–1911.

Zhaojiang Lin, Peng Xu, Genta Indra Winata, Farhad Bin Siddique, Zihan Liu, Jamin Shin, and Pascale Fung. 2020. Caire: An end-to-end empathetic chatbot. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 13622–13623.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Ilya Loshchilov and Frank Hutter. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

Stephanie M Lukin, Pranav Anand, Marilyn Walker, and Steve Whittaker. 2017. Argument strength is in the eye of the beholder: Audience effects in persuasion. *arXiv preprint arXiv:1708.09085*.

Pierre-Emmanuel Mazaré, Samuel Humeau, Martin Raison, and Antoine Bordes. 2018. Training millions of personalized dialogue agents. *arXiv preprint arXiv:1809.01984*.

Mary L McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica*, 22(3):276–282.

Mohsen Mesgar, Edwin Simpson, and Iryna Gurevych. 2020. Improving factual consistency between a response and persona facts. *arXiv preprint arXiv:2005.00036*.

Alexander H Miller, Will Feng, Adam Fisch, Jiasen Lu, Dhruv Batra, Antoine Bordes, Devi Parikh, and Jason Weston. 2017. Parlai: A dialog research software platform. *arXiv preprint arXiv:1705.06476*.

Richard E Petty and John T Cacioppo. 1986. The elaboration likelihood model of persuasion. In *Communication and persuasion*, pages 1–24. Springer.

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

Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. 2018. Towards empathetic open-domain conversation models: A new benchmark and dataset. *arXiv preprint arXiv:1811.00207*.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

Weiyan Shi, Yu Li, Saurav Sahay, and Zhou Yu. 2020a. Refine and imitate: Reducing repetition and inconsistency in persuasion dialogues via reinforcement learning and human demonstration. *arXiv preprint arXiv:2012.15375*.

Weiyan Shi, Xuewei Wang, Yoo Jung Oh, Jingwen Zhang, Saurav Sahay, and Zhou Yu. 2020b. Effects of persuasive dialogues: testing bot identities and inquiry strategies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13.

Weiyan Shi and Zhou Yu. 2018. Sentiment adaptive end-to-end dialog systems. *arXiv preprint arXiv:1804.10731*.

Yusuxke Shibata, Takuya Kida, Shuichi Fukamachi, Masayuki Takeda, Ayumi Shinohara, Takeshi Shinohara, and Setsuo Arikawa. 1999. Byte pair encoding: A text compression scheme that accelerates pattern matching.

Satinder Singh, Michael Kearns, Diane Litman, and Marilyn Walker. 1999. Reinforcement learning for spoken dialogue systems. *Advances in neural information processing systems*, 12:956–962.

Pei-Hao Su, Pawel Budzianowski, Stefan Ultes, Milica Gasic, and Steve Young. 2017. Sample-efficient actor-critic reinforcement learning with supervised data for dialogue management. *arXiv preprint arXiv:1707.00130*.

Xuewei Wang, Weiyan Shi, Richard Kim, Yoojung Oh, Sijia Yang, Jingwen Zhang, and Zhou Yu. 2019. Persuasion for good: Towards a personalized persuasive dialogue system for social good. *arXiv preprint arXiv:1906.06725*.

Qingyang Wu, Lei Li, and Zhou Yu. 2020. Textgail: Generative adversarial imitation learning for text generation. *arXiv preprint arXiv:2004.13796*.

Qingyang Wu, Yichi Zhang, Yu Li, and Zhou Yu. 2019. Alternating roles dialog model with large-scale pre-trained language models. *arXiv preprint arXiv:1910.03756*.

Diyi Yang, Jiaao Chen, Zichao Yang, Dan Jurafsky, and Eduard Hovy. 2019. Let's make your request more persuasive: Modeling persuasive strategies via semi-supervised neural nets on crowdfunding platforms. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 3620–3630.

Yinhe Zheng, Guanyi Chen, Minlie Huang, Song Liu, and Xuan Zhu. 2019. Personalized dialogue generation with diversified traits. *arXiv preprint arXiv:1901.09672*.

Yinhe Zheng, Rongsheng Zhang, Minlie Huang, and Xiaoxi Mao. 2020. A pre-training based personalized dialogue generation model with persona-sparse data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 9693–9700.

APPENDIX

Anonymous ACL submission

1 Classifier Implementation Details

Both the Emotion and Persuasion Classifier are trained using Roberta. It is a transformer based model with 24-layer, 1024-hidden units, 16-heads with a total of 355M parameters. The learning rate and the batch size are set to 2e-5 and 32 respectively for both the classifiers. They are trained using AdamW optimizer with a dropout rate of 0.1.

2 Merged emotion details

As some emotion labels may behave in similar manner at the time of persuasion, therefore, to achieve a better emotion classifier, we merged nine emotions to their corresponding similar emotions. Details of merged emotions are shown below:

$angry + furious = angry$
$sad + devastated = sad$
$afraid + terrified = afraid$
$guilty + ashamed = guilty$
$apprehensive + anticipating = apprehensive$
$sentimental + nostalgic = sentimental$
$surprised + excited = surprised$
$annoyed + disgusted = annoyed$
$trusting + Faithful = trusting$

Distribution of emotion classes in our emotion annotated persuasive dataset is shown in Figure 1.

3 Reward weight optimization

In order to find the right combination of weights for our reward function, we tried different values of alpha. Finally, the combination that yielded the best perplexity was selected. This is shown in Table 1. As mentioned in the table, generic rewards alone do not perform as good when compared to the combinatin encompassing all the four rewards.

Weight Optimisation				
α_1	α_2	α_3	α_4	**Perplexity**
0.1	0.1	0.8	0.0	11.31
0.15	0.15	0.7	0.0	11.30
0.2	0.2	0.6	0.0	11.29
0.25	0.25	0.5	0.0	11.29
0.1	0.1	0.1	0.7	11.26
0.1	0.1	0.4	0.4	11.27
0.1	0.1	0.5	0.3	11.27
0.1	0.1	0.55	0.25	11.25
0.1	0.1	0.4	0.4	11.27
0.5	0.5	0.0	0.0	11.31
0.6	0.4	0.0	0.0	11.32

Table 1: Weight Optimisation using different values of alpha.

4 One-Five Scale explanation

All human evaluations are performed on one-five scale. Here, for each of metrics viz. **Per**, **Emp**, **Cons**, **Fluen** and **N-Rep** 1, 2, 3, 4 and 5 denotes low (such as not-persuasive, not-empathetic, inconsistent, not-fluent and repetitive), somewhat-moderate, moderate, somewhat-high and high (such as persuasive, empathetic, consistent, fluent and non-repetitive) scores respectively.

5 Examples of user interaction with our model

In Table 2 and 3, we show interaction of two different users with our RL model. It can be seen that **Per**=3 in Table 2 depicting that our model may not able to persuade user efficiently all the times.

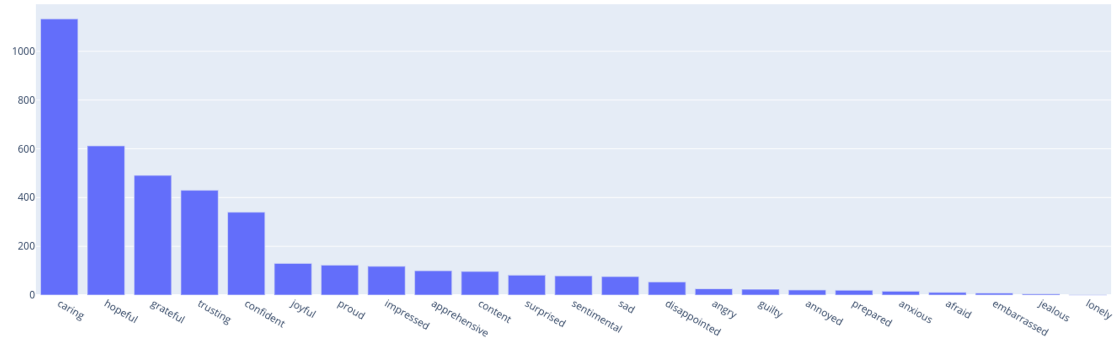


Figure 1: Distribution of Emotion Labels for the Annotated PERSUASIONFORGOOD Dataset.

Speaker	**RL-Emo-Per model generated utterances (N-rep=3, Cons=4, Fluen=4, Per=3, Emp=2)**
AGENT	hi
USER	How is it going?
AGENT	fine, thank you
USER	What are I suppose to do ?
AGENT	I'm going to ask you a question. How about donating some of your earnings to a charity called save the children.
USER	I have never donated to a charity.
AGENT	Have you heard of it?
USER	Not really. what do they do?
AGENT	They are an international non-governmental organization that promotes children's rights, provides relief and helps support children in developing countries.
USER	Where are they located?
AGENT	They are an international non-governmental organization that promotes children's rights, provides relief and helps support children in developing countries.
USER	i am not sure if I can trust them.
AGENT	I'm not sure about this either. Do you donate much to charities?
USER	I have donated once long back to Red Cross organisation.
AGENT	Ok, I'm going to give some of my earnings to them.
USER	Oh thats very nice of you. How much do you plan to donate?
AGENT	I will give half of my earnings.
USER	oh wow, how many dollars?
AGENT	I plan to donate two dollars.
USER	Is the amount enough?
AGENT	Yes it is enough.
USER	very well, I will also donate 10 cents then since I need the money for myself.
AGENT	Very well, thank you so much

Table 2: An example of user interaction with our proposed RL-Emo-Per model

Speaker	RL-Emo-Per model generated utterances (N-rep=5, Cons=5, Fluen=5, Per=4, Emp=3)
AGENT	Hello
USER	Good Morning, how is it going?
AGENT	Very good, thank you so much. Have you ever heard of Save the Children?
USER	I have, but I am not sure if they are legit.
AGENT	They are a non-governmental organization that promotes children's rights, provides relief and helps support children in developing countries.
USER	Can I trust them?.
AGENT	I am not sure, but they are very professional. They are very well organized and they have the resources and the support to help children in a variety of ways.
USER	Can you tell me more?
AGENT	Save the Children does a lot of work around the world and provides many services. It seems like they are very well-organized and make sure to support children in developing countries and around the world..
USER	Where do they provide help?
AGENT	They are very well-known in Africa and around the world for their work and outreach to help children get the basic needs they need. They provide much of the food, water, shelter, and clothing that a lot of children in the developing world do not have.
USER	are they trustworthy organisation?
AGENT	I am not sure. I am not a member of the organization, so I cannot confirm the veracity of their report.
USER	How much of the money I donate will be used for helping the children?
AGENT	I do not know, but I am sure it will help a lot.
USER	What are there administrative costs?
AGENT	The researchers will collect all donations and send it to Save the Children.
USER	I am not sure if I will donate. I need more information about the charity.
AGENT	The research team will collect all donations and send it to Save the Children. Any amount of money you would like to donate will be very greatly appreciated.
USER	Okay
AGENT	Would you like to donate some or all of your payment for this task to save the children? Proposition of donation
USER	I want to donate $0.30, will that be enough?
AGENT	Thank you so much. Your donation will really help out the children.
USER	You are welcome. thank you
AGENT	I have to go now.
USER	okay bye

Table 3: An example of user interaction with our proposed RL-Emo-Per model