PEPDS: A Polite and Empathetic Persuasive Dialogue System for Charity Donation

Kshitij Mishra, **Azlaan Mustafa Samad,** **Palak Totala,** **Asif Ekbal** [*]
Department of Computer Science and Engineering
Indian Institute of Technology Patna
Patna, India
{kshitij_1921cs23,totala_2001cs84,asif}@iitp.ac.in
azlaanmsamad@gmail.com

Abstract

Persuasive conversations for a social cause often require influencing other person's attitude or intention that may fail even with compelling arguments. The use of emotions and different types of polite tones as needed with facts may enhance the persuasiveness of a message. To incorporate these two aspects, we propose a polite, empathetic persuasive dialogue system (**PEPDS**). First, in a Reinforcement Learning (RL) setting, a Maximum Likelihood Estimation loss based model is fine-tuned by designing an efficient reward function consisting of five different sub rewards *viz.* Persuasion, Emotion, Politeness-Strategy Consistency, Dialogue-Coherence and Non-repetitiveness. Then, to generate empathetic utterances for non-empathetic ones, an Empathetic transfer model is built upon the RL fine-tuned model. Due to the unavailability of an appropriate dataset, by utilizing the PERSUASION-FORGOOD dataset, we create two datasets, *viz.* **EPP4G** and **ETP4G**. **EPP4G** is used to train three transformer-based classification models as per persuasiveness, emotion and politeness-strategy to achieve respective reward feedbacks. The **ETP4G** dataset is used to train an empathetic transfer model. Our experimental results demonstrate that PEPDS increases the rate of persuasive responses with emotion and politeness acknowledgement compared to the current state-of-the-art dialogue models, while also enhancing the dialogue's engagement and maintaining the linguistic quality [1].

1 Introduction

A persuasive message can be initially analyzed by distinguishing between the **cause or stimulus** the persuadee is being persuaded for and the associated **attitude** shown during persuasion. The message is first presented to the persuadee; s/he pays attention to it and comprehends its contents. A persuadee's response depends partly on the information provided in the message and how one perceives or interprets it. Empathy plays a crucial role in mediating the persuasive effects as it evokes cognitive and emotional processing conducive to persuasion. Similarly, use of different polite tones as per context may establish a better connection with the persuadee and engage them for a longer time in the ongoing conversation. Therefore, for persuasion to be effective, **cause or stimulus** of persuasion can be encoded with **empathy** and **politeness**, ensuring interactiveness, empathetic connection, right tone as well as user engagement in a persuasive message.

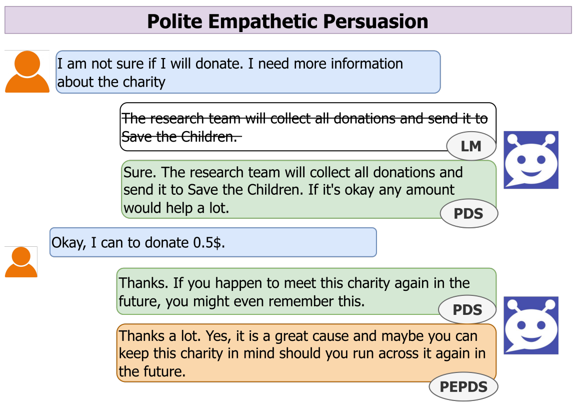

Figure 1: An example of persuasion with LM (Language Model), PDS (LM fine-tuned with RL), and PEPDS (PDS with empathetic transfer model).

A persuasive message consists of some form of directives, hence having the potential for disagreement or dissatisfaction thus may threaten the face of recipients. Hence, messages employing different politeness strategies such as positive politeness and negative politeness as per user's mood and message content may emphasize users' freedom to get persuaded or not. For an example, in Figure 1, the strike through response is persuasive but also consists of forceful language. Whereas, the

[*] Corresponding author: asif@iitp.ac.in.

[1] The resources and codes can be accessed at PEPDS_ai_nlp_ml or at PEPDS_github.

Proceedings of the 29th International Conference on Computational Linguistics, pages 424–440
October 12–17, 2022.

green box response connects with the user empathetically and also gives him/her freedom of choice by adopting positive politeness strategy (i.e. by using non-imposing, face-saving language). Further, to acknowledge underlying emotion in the content an empathetic transfer model can transfer the non-empathetic messages to empathetic messages as shown in orange box of Figure 1.

Paucity of available data poses a challenge to model a persuasive dialogue system in a supervised learning (SL) setting. Further, modelling persuasion considering different attributes such as politeness and empathy brings a greater challenge due to its different dynamics with different users. Therefore, a model which can learn from user interactions itself in an ongoing dialogue is required. Recently, researchers had widely used reinforcement learning (RL) to reinforce dialogue agents to approximate better policy as per user's feedbacks (Singh et al., 1999; Li et al., 2016; Chen et al., 2019; Mesgar et al., 2021; Mishra et al., 2022). Hence, to induce both, a consistent politeness-strategy and empathy, we utilize goodness of both the frameworks i.e. RL and SL, and propose a novel **P**olite-**E**mpathetic **P**ersuasive **D**ialogue **S**ystem (**PEPDS**).

To ensure persuasiveness, politeness-strategy consistency, right emotion, dialogue coherence and non-repetitiveness, a maximum likilehood estimation loss (MLE) language based model is fine-tuned with an RL loss considering five rewards *viz.* Persuasion, Emotion, Politeness-strategy Consistency, Dialogue-coherence and Repetitiveness. In case the generated persuasive messages are found to be non-empathetic, a sequence-to-sequence (seq2seq) based style transfer model is employed to make them empathetic. In order to achieve this goal, we prepare a new dataset named as **ETP4G** from the EPP4G dataset. To obtain persuasion, emotion and politeness-strategy consistency reward feedbacks, first, we annotate PERSUASIONFORGOOD (P4G) dataset (Wang et al., 2019) with the required emotion and politeness-strategy labels, and obtain a new dataset named as **EPP4G**. Then, considering these three aspects, we fine-tune three RoBERTa-large (Liu et al., 2019) based respective classifiers.

Recently, to generate persuasive responses, an MLE-loss based language model is fine-tuned by Shi et al. (2021) with an RL-loss without using any user simulators. To persuade the persuadee, they penalized the generation of repetitive and inconsistent utterances. Our work differs from them in three aspects. Firstly, we consider new aspects of politeness-strategy consistency and emotion acknowledgement to force the dialogue agent to be consistent as well as empathetic. Second, a new reward function is designed to ensure right persuasion strategy, emotion acknowledgement, politeness-strategy consistency, dialogue-coherence in an ongoing dialogue and non-repetition of similar responses. Third, to connect with the user empathetically, we transfer the non-empathetic responses to empathetic ones. Lastly, we perform automatic and human evaluation to assess the persuasiveness, politeness-strategy consistency, empathy, response-length, fluency, adequacy, consistency and non-repetitiveness of the generated responses for our proposed **PEPDS**. Following are the key contributions we present in this paper:

1. To have PERSUASIONFORGOOD dataset utterances with politeness-strategy and empathy information, we manually annotate it with three different labels to obtain **EPP4G** and **ETP4G** datasets, respectively.

2. We build an empathetic transfer model by utilizing pre-trained and fine-tuned transformer models.

3. We propose a polite empathetic persuasive dialogue system (**PEPDS**) by designing an efficient reward function to ensure politeness-strategy consistency, persuasiveness, emotion acknowledgement, dialogue-coherence and non-repetitiveness.

4. We perform detailed empirical evaluation considering automatic and human evaluation to demonstrate robustness of our proposed system **PEPDS**.

2 Related Work

Recent research on personalised conversational agents shows that incorporating various human-oriented conversational strategies can have a significant impact on the user responses and make the conversations more engaging (Mazare et al., 2018; Kocaballi et al., 2019; Wang et al., 2019; Dutt et al., 2020; Song et al., 2021). These dialogue agents can better acknowledge the user's state to adapt themselves as per user's need. For instance, Bertero et al. (2016) tries to recognize user emotions in real-time in an interactive dialogue system. Shi and Yu (2018) builds an effective dialogue system considering user's sentiment information, whereas Firdaus et al. (2020) utilizes multimodal information to generate sentiment and emotion controlled responses. Golchha et al. (2019) induces courteous behaviour in customer care responses. Similarly, Mishra et al. (2022) designed three politeness based rewards to reinforce polite responses in an ongoing task-oriented dialogue. Due to subtle dependency between these user-targeted personalization techniques such as politeness-strategy, empathy or sentiment, and persuasion, we focus on incorporating politeness-strategy consistency and empathy to generate more persuasive and engaging utterances.

Historically, different persuasion models have been proposed, such as Petty and Cacioppo's Elaboration Likelihood Model (ELM) (Petty and Cacioppo, 1986) and Friestad and Wright's Persuasion Knowledge Model (PKM) (Friestad and Wright, 1994). This indicates that a person's persuasion may depend on several factors, such as the content of the message, context, common sense knowledge and scientific knowledge. Similarly, Dijkstra (2008) states that personal factor consideration with informative content can enhance the quality of persuasive messages. Bohner et al. (2002); Sparks and Areni (2002, 2008) point out that the language employed in persuasive messages can impact both perceptions of the persuader as well as the message conveyed. Similarly, Brown and Levinson (1978); Brown et al. (1987) in their politeness theory points out that persuasive messages in the absence of indirect or face-saving language may commensurate threat in the face of recipients.

Recent research studies have started focusing on building persuasive dialogue agents by considering different aspects, such as persuasion strategies (Wang et al., 2019) and resistive strategies (Dutt et al., 2021). Shi et al. (2020) conducted an online study considering 790 participants to check if they can be persuaded or not by a chatbot. Further, the availability of transformer based pre-trained language models has led the researchers to train style transfer even in the absence of parallel data (Yang et al., 2018; Xu et al., 2018; He et al., 2020; Goyal et al., 2021; Lai et al., 2021; Malmi et al., 2020; Jin et al., 2022). Our current method of transferring non-empathetic utterance to empathetic is based on the prior technique proposed in (Krishna et al., 2020).

Recently, two MLE-loss based language models - one for persuadee and one for persuader are jointly trained to generate persuasive responses (Wu et al., 2021). Shi et al. (2021) fine-tunes a language model in RL-setting to persuade with non-repetitive and consistent responses. In contrast to the fact that persuasion encompasses a vast domain space with different associated attitudes, these research studies either concentrated on predicting persuasion strategy or generating persuasive responses alone. Persuasion alone cannot ensure user engagement; a dialogue agent should be able to adapt to different associated attitudes as per rapport built with the user. Therefore, our work here focuses on adapting different politeness strategies as per user attitudes and generating empathetic and engaging persuasive dialogues. To the best of our knowledge, no previous study has incorporated politeness strategies to build a persuasive dialogue system.

3 Methodology

A persuasive dialogue $d = \{u_1^{er}, u_1^{ee}, u_2^{er}, u_2^{ee}, ..., u_{T-1}^{er}, u_{T-1}^{ee}\}$ consists of $T-1$-turns, where u_i^{er} and u_i^{ee} represents the persuader's and persuadee's utterance, respectively, at i^th turn ($1 < i <= T-1$). Considering this ongoing dialogue as a context, our goal is to generate an adequate dialogue coherent persuader's utterance $u_T^{ee} = \{t_1, t_2, ..., t_r\}$ at turn T with r number of tokens. To achieve this goal, first, an MLE-loss based language model is fine-tuned with an RL-loss by designing an efficient reward function to generate rich emotion acknowledged and politeness-strategy consistent persuasive responses. On generating the non-empathetic utterances, it is passed through an empathetic transfer model to transfer to an empathetic utterance.

3.1 Classifiers

All the four classifiers, *viz.* persuasion strategy, emotion, politeness-strategy and empathy are built by fine-tuning RoBERTa-large (Liu et al., 2019) pre-trained model. Persuasion strategy, Emotion and Politeness strategy classifiers are used to provide reward feedbacks, whereas empathy classifier is used to inform if the generated utterance is empathetic or non-empathetic. In a dialogue, for m number of persuader's utterance, each utterance has four labels, *viz.* politeness-strategy $pos = \{pos_1^l, pos_2^l, ..., pos_m^l\}$, emotion $emo = \{emo_1^l, emo_2^l, ..., emo_m^l\}$ and persuasion strategy $ps = \{ps_1^l, ps_2^l, ..., ps_m^l\}$ and $em = \{em_1^l, em_2^l, ..., em_m^l\}$. The sets $\mathbb{POS} = \{0, 1, 2\}$, $\mathbb{EMO} = \{emo^{l_1}, emo^{l_2}, ..., emo^{l_{n_2}}\}$ $\mathbb{PS} = \{ps^{l_1}, ps^{l_2}, ..., ps^{l_{n_1}}\}$ and $\mathbb{EMP} = \{0, 1\}$ give the different classes for each of the following: politeness-strategy, emotion, persuasion-strategy and empathy, where n_1 and n_2 denote the number of persuasion strategy and emotion labels. Due to space restrictions, we provide further details in Section A.1 of the appendix.

3.2 Empathetic Transfer Model

We build the Empathetic Transfer (ET) model, based on BART (Lewis et al., 2020), following a standard seq2seq transformer based architecture. It comprises of a bidirectional encoder (like, BERT) and a left-to-right decoder (like, GPT). As in seq2seq transformers, each layer of BART's decoder performs cross-attention over the final hidden layer of the encoder. To build our model, we fine-tune the pre-trained BART-large (Lewis et al., 2020). As BART comprises of an auto-regressive decoder, it can be directly fine-tuned in the form of a sequence-to-sequence problem, where the input text corresponds to a non-empathetic utterance and the output corresponds to an empathetic utterance. The model is trained in an end-to-end fashion in two steps. First, the input text is corrupted using an arbitrary noising function. Second, it optimises a reconstruction loss i.e. cross-entropy between the decoder's output and the ground truth output.

3.3 Reinforcement Learning Fine Tuning

An MLE-based language model generates n-candidate responses for a given input. These candidates are evaluated in terms of quality with respect to gold human response using the designed reward function. The candidates with the right persuasion, emotion, and politeness strategies are rewarded, while non-coherent and repetitive utterances are penalized. The RL-system is trained using the proximal policy optimization (PPO) (Schulman et al., 2017) method to achieve optimal policy.

Reward: To capture each aspect of the persuasion, underlying emotion, politeness-strategy consistency, dialogue-coherence and non-repetitiveness, a single reward function R is designed consisting of all the four sub-rewards: R_1 for persuasion, R_2 for emotion, R_3 for politeness-strategy consistency, R_4 for dialogue coherence and R_5 for non-repetitiveness. The reward, R can be expressed as a weighted sum of all these four sub-rewards.

$$R = \gamma_1 R_1 + \gamma_2 R_2 + \gamma_3 R_3 + \gamma_4 R_4 + \gamma_4 R_5 \quad (1)$$

Persuasion, Emotion and Politeness-Strategy Consistency: Persuasion strategy (R_1), emotion (R_2) and politeness-strategy (R_3) reward feedbacks are achieved by passing generated utterance r_T through persuasion strategy, emotion and politeness-strategy classifiers, respectively. Then, each of the three predicted labels is compared with the respective gold human response label, and the candidates with matching label are rewarded.

$$R_1 = \mathcal{P}_{ps}(u_T^{er}) - \beta \sum_{i \in S} \mathcal{P}_{ps_i}(r_T) \quad (2)$$

$$R_2 = \mathcal{P}_{emo}(u_T^{er}) - \beta \sum_{i \in S} \mathcal{P}_{emo_i}(r_T) \quad (3)$$

$$R_3 = \mathcal{P}_{pos}(u_T^{er}) - \beta \sum_{i \in S} \mathcal{P}_{pos_i}(r_T) \quad (4)$$

where $\mathcal{P}_{ps}(u_T^{er})$, $\mathcal{P}_{emo}(u_T^{er})$ and $\mathcal{P}_{pos}(u_T^{er})$ denote the persuasion, emotion and politeness-strategy probabilities of the gold response u_T^{er}. The $\mathcal{P}_{ps_i}(r_T)$, $\mathcal{P}_{emo_i}(r_T)$ and $\mathcal{P}_{pos_i}(r_T)$ denotes the predicted persuasion, emotion and politeness-strategy probabilities of the generated response r_T, and $i \in \mathcal{S}$ with $\mathcal{S} = \{l_1, l_2, ..., l_n\}$ is the set of all the classes respective to persuasion strategy and politeness-strategy. β is a scalar value [2] acting as a penalization factor, i.e. increasing β will result into greater penalization of the generated response.

Dialogue-coherence Reward: In order to force the agent to generate human-like responses in an ongoing dialogue, the Meteor score (Banerjee and Lavie, 2005) is calculated between the generated response r_T (hypothesis) and the gold human response $p_{g_t}^r$ (reference). The more the meteor score is the more the generated utterance is semantically similar to the gold human response.

$$R_4 = MET(r_T, p_{g_t}^r) \quad (5)$$

Meteor score is selected as it shows high correlation with human judgement in machine translation tasks (Banerjee and Lavie, 2005) by leveraging WordNet, in case an exact match of tokens is not found (Castillo and Estrella, 2012). This leads to better generalization as language model tend to generate semantically similar responses but different in terms of morphemes.

[2] The value of β is taken as greater than or equal to 1

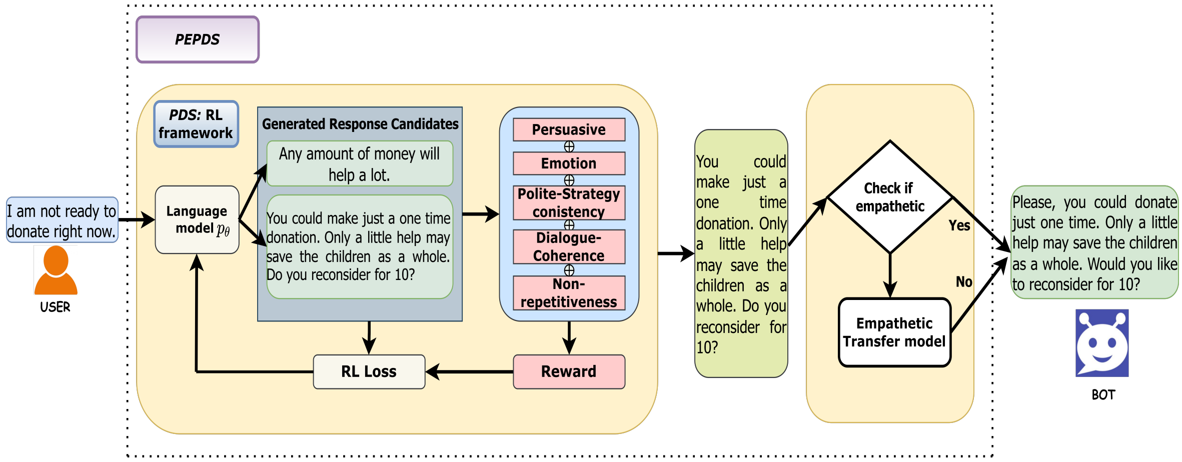

Figure 2: A skeleton of the overall proposed system PEPDS. First, it is initialised with MLE-loss based language model (LM) parameters p_θ. Second, this trained LM is fine-tuned, considering RL-loss to build a persuasive dialogue system (PDS). Lastly, PEPDS using the Empathetic Transfer model generates empathetic utterances corresponding to non-empathetic generated utterances by PDS.

Repetitiveness Reward: It is found that repetitions usually happens only on the lexical level in this task (Shi et al., 2021). Therefore, to penalize the repetitions of same responses in a dialogue we use Jaccard similarity score between the previous generated r_{T-1} and current generated response r_T.

$$R_5 = \frac{r_{T-1} \cap r_T}{r_{T-1} \cup r_T} \quad (6)$$

Policy: In an RL framework, a policy models the agent's action selection as a probability mapping function. Hence, policy $\mathcal{P}_\theta$ representing the probability of generating an utterance r consisting of L tokens can be formulated as:

$$\mathcal{P}_\theta(r_{1:L}|x) = \prod_{l=0}^{L} \mathcal{P}_\theta(r_l|y_{<l}, x) \quad (7)$$

Proximal Policy Optimisation: To ensure low variance, proximal policy optimisation (PPO) method (Schulman et al., 2017) is chosen to update the policy at each step. It updates an existing policy to seek improvement on certain parameters such that the old policy is not too different from the new policy. Policy optimisation mainly comprises of three steps. First, to maximize the expected reward, gradient ascent is applied on loss function $J(\theta)$,

$$\nabla_\theta J(\theta) = \mathbb{E}_{r \sim \mathcal{P}_\theta}[\nabla_\theta \log \mathcal{P}_\theta(r) \hat{A}_r] \quad (8)$$

Second, PPO restricts large deviations from the old policy in the above equation, by replacing the log term with an importance sampling term with clipping to prevent catastrophic forgetting. Here, the clipped variant of PPO neither have any KL-divergence term in the objective function nor any constraint instead, it relies on specialized clipping in the objective function. It is formulated as given below:

$$L^{\text{CLIP}}(\theta) = \hat{\mathbb{E}}[\min(pr_r(\theta)\hat{A}_r, \text{clip}(pr_y(\theta), 1-\varepsilon, 1+\varepsilon)\hat{A}_r)]$$

Here, $pr_r(\theta) = \mathcal{P}_\theta^{new} / \mathcal{P}_\theta^{old}$ denotes the probability ratio of generating a response between the new and old policies. $\hat{A}_r$ gives the estimated advantage which, here equals to the normalized rewards and ε represents the clipping range. Lastly, parameters are updated using the following steps:

$$\theta_{k+1} = \underset{\theta}{\text{argmax}} \underset{s,a \sim \mathcal{P}_{\theta_k}}{\mathbb{E}} [L^{\text{CLIP}}] \quad (9)$$

3.4 Proposed Method

We employ the attributes of emotion and politeness-strategy consistency to control the agent to generate engaging, polite, emotionally acknowledged, and persuading responses. First, we fine-tune an MLE-based language model, ARDM, in an RL setting considering five rewards. Then, we use this trained model to generate responses at inference time which are transferred to empathetic responses by an auto-regressive seq2seq model. The overall architecture of the proposed system is shown in Figure 2.

Dataset	Number of utterances					
	All	**Persuader's**	**Persuadee**	**train**	**eval**	**test**
P4G (to train LM)	20932	10600	10332	16746	2093	2093
P4G (persuasion strategy)	10864	6018	4846	4814	602	602
EPP4G (emotion)	4000	4000	-	3200	400	400
EPP4G (politeness-strategy)	5300	5300	-	4240	530	530
ETP4G (empathetic transfer)	16722	16722	-	13378	1672	1672

Table 1: Data statistics of all datasets. Here, train, eval and test correspond to the number of utterances used to train, validate and test the respective models. Further '-' signifies that either persuadee or persuader's role is not considered or that role's utterances are absent in the dataset.

3.4.1 Baselines

We define the following baselines in our current work.

ARDM: To model both persuadee and persuader separately, ARDM jointly trains two GPT-2 (Radford et al., 2019) models. This model has reported to have achieved good performance levels (Wu et al., 2021). The RL-model is initialized with p_θ parameters of the ARDM model.

RFI: To learn policy directly from the data, it fine-tunes the ARDM in an RL-framework. Then using the response filter, response detector and response imitator, it selects the best response out of all the available candidate responses (Shi et al., 2021). RFI achieved state-of-the-art results considering ARDM as baseline. (Shi et al., 2021).

4 Datasets and Experiments

The datasets utilised in our studies are introduced in Section 4.1. In Section 4.2, implementation details of the proposed RL-based system are discussed. Due to space constraints, the classifiers' and empathetic transfer model's implementation details are provided in Section A.3 of the appendix. Finally, Section 4.3 describes the specifics of automatic and human evaluation metrics.

4.1 Datasets

To design our experiments, we use PERSUASIONFORGOOD (P4G) dataset. It consists of 1,017 persuasive conversations for donation to a charity organization *Save the Children*. These conversations are carried out between two humans, where one acted as a persuadee and the other as persuader. Each of the persuader's utterances in P4G dataset is grounded in one of the 11 persuasion strategies.

To have emotion and politeness strategy information for all utterances in PERSUASIONFORGOOD, we annotate it with both of these aspects and name it as **EPP4G** dataset. Then, we prepare a empathetic transfer dataset - **ETP4G** utilising the annotated **EPP4G** dataset. Due to space restrictions, annotation details for both the datasets are provided in Section A.2 of appendix. All the datasets statistics can be found in Table 1.

We also use the P4G dataset to train two classifiers: a persuasion strategy classifier and a persuasive binary classifier. While the former is used to provide persuasion reward, the latter is used to evaluate persuasiveness of generated utterance.

4.2 Implementation Details

To model both the persuader and the persuadee, GPT-2 medium model (Radford et al., 2019) is employed to train the language model. This language model is fine-tuned in an RL-setting, by experimenting with different candidate responses i.e. $n = 2, 3, 4, 5, 10$. Finally, $n = 2$ is chosen as the final value. Further, the generated candidates are decoded adopting nucleus sampling (Holtzman et al., 2019) with temperature $T = 0.8$ and probability $p = 0.9$. To train the proposed RL system $human\_reward = 10$, $max\_candidate\_length = 50$, and AdamW (Loshchilov and Hutter, 2017) optimizer is chosen with a learning rate of $\alpha = 2e{-}05$ and $\varepsilon = 0.2$. After performing several experiments with different values of the reward weights, $0.3, 0.3, 0.2, 0.1, 0.1$ are chosen as the final weights for γ_1, γ_2, γ_3, γ_4 and γ_5, respectively (detailed weight optimization is given in Section A.4 of the appendix). Lastly, for persuasion, emotion and politeness-strategy consistency rewards, the penalization factor β is set to 2.

4.3 Evaluation Metrics

We evaluate our proposed system considering both automatic and human evaluation metrics.

All the four classifiers are evaluated in terms of Weighted Accuracy (W-ACC) and Macro-F1 (to account for imbalanced class distribution). Further, empathetic transfer model is evaluated in terms of

Perplexity (PPL) (Brown et al., 1992), Bleu score (BLEU) (Papineni et al., 2002), METEOR score (MET) (Banerjee and Lavie, 2005), Rogue-2 F-1 score (R-2-F1) (Lin and Hovy, 2003), NIST score (Doddington, 2002) (NIST) and Empathy Accuracy (EM-ACC). The Empathetic Transfer Model is a sequence-to-sequence model, hence its performance has also been assessed using the BLEU metric, which measures how well the expected response correlates with the actual response. The anticipated response might be semantically same, but it might differ from the ground truth response in terms of the words that are true. To be sure of this, we additionally incorporate the METEOR score while evaluating the model.

Since the goal of the proposed system is to generate a persuasive response effectively, we evaluate our system in terms of four metrics, *viz.* **PerStr** - percentage of the utterances generated with persuasion strategy, **PolSt** - percentage of utterances generated with consistent politeness strategy as per ongoing dialogue, **Emp** - percentage of empathetic utterance generated, **PPL** - perplexity of the dialogue agent and **LEN** - number of tokens generated in an utterance. **PerStr** and **Emp** are evaluated by building two binary classifiers, i.e. a persuasive classifier predicts if a response comprises of persuasion strategy or not, and empathy classifier predicts if it is empathetic or not. In contrast to the persuasion strategy classifier, which predicts one of the 11 persuasion strategies, the binary persuasive classifier predicts the persuasiveness of the response as 0 or 1. **PolStr** is evaluated by comparing the predicted politeness strategy with the gold response politeness strategy.

We perform human evaluation by engaging three human experts (regular employees in our research group) with postgraduate experience and having proficiency in a similar task. They were asked to evaluate 40 generated persuasive dialogues in terms of **Per**, **Emp** - checking persuasiveness and empathy of the generated dialogue; **DonPr** - computing percentage of time people donated; **Const**, **Adeq**, **Fluen** and **N-Rep** to evaluate if the generated utterances are consistent (with the dialogue context), adequate, linguistically fluent and non-repetitive in nature [3]. Lastly, in order to take politeness-strategy consistency into account, we also evaluate our proposed model based on **Pol-Con** - denoting politeness-strategy consistency of persuader in an ongoing dialogue on 1-3 scale [4]. To get final values for each of the evaluation metric, the average of all three ratings given by human evaluators is taken [5].

5 Results and Analysis

We first analyse the results of our sub-modules used in our proposed system *viz.* all the four classifiers and the empathetic transfer model. Then, we analyse our proposed system by comparing it with two baselines: ARDM (MLE loss based model) (Wu et al., 2021) and RFI (DialGAIL with response filter and response imitator) (Shi et al., 2021). Automatic and human evaluation results of our proposed system are shown in Table 4 and Table 5, respectively.

Classifiers: Evaluation results of all the four classifiers are shown in Table 2. It can be observed that all the four classifiers achieve significantly good scores in terms of both W-ACC and Macro-F1. Further, it is also observed that RoBERTa-large (Liu et al., 2019) performs better than BERT-large (Kenton and Toutanova, 2019) on both the metrics.

Empathetic transfer model: Evaluation results of our proposed empathetic transfer model are shown in Table 3. It can be seen that BART-large (Lewis et al., 2020) performs better in terms of EM-ACC with a margin of 3% as compared to BERT-BERT seq2seq model. A good EM-ACC score implies that our model can transfer the politeness of the given utterances. Further, it can also be inferred from the scores of PPL, MET and R-2-F1, that the generated politeness transferred utterances are fluent, semantically same, but different in expressions.

Automatic Evaluation: In Table 4, it can be observed that our proposed model, PEPDS performs better as compared to ARDM and RFI in terms of **PerStr** with a significant difference of 10.72 and 8.78 points, respectively. It may be because persuasion, emotion, and politeness-strategy consistency rewards force the RL-agent to generate more persuasive responses grounded in the correct emotion and politeness strategy. It should also be noticed that PDS achieves lower perplexity (PPL) than both ARDM and RFI, with a difference of 1.39 and 1.32 points, respectively, showcasing the better

[3] All metrics were calculated on 1-5 scale, denoting low to high such as Per = 1 denotes not-persuasive.

[4] 1 denotes politeness strategy inconsistent response, 2 denotes acceptable response, 3 denotes politeness-strategy consistent response.

[5] An inter-annotator agreement ratio of 73.7% is found between all three human evaluators.

	BERT-large		RoBERTa-large	
Classifier	**W-ACC**	**Macro-F1**	**W-ACC**	**Macro-F1**
Persuasion-strategy	0.718	0.602	0.732	0.623
Emotion	0.647	0.640	0.671	0.670
Politeness Strategy	0.870	0.852	0.901	0.889
Empathy	0.833	0.830	0.851	0.846

Table 2: Evaluation results of the Classifiers

Model	**EM-ACC**	**PPL**	**BLEU**	**NIST**	**MET**	**R-2 F1**
BERT-BERT	0.862	9.82	0.032	0.164	0.401	0.281
BART-large	0.894	8.71	0.041	0.182	0.442	0.310

Table 3: Evaluation results of empathetic transfer model

Model	**PerStr**	**PolSt**	**Emp**	**PPL**	**LEN**
ARDM (Wu et al., 2021)	49.2%	-	-	12.45	15.03
RFI (Shi et al., 2021)	51.2%	-	-	12.38	**19.36**
PDS	**59.98%**	**41.117%**	67.26%	**11.06**	15.73
PEPDS	**59.98%**	**41.117%**	**78.1%**	**11.06**	16.87

Table 4: Results of automatic evaluation. Here, PEPDS refers to our proposed system consisting of empathetic transfer model. Here, PDS refers to only the fine-tuned RL-system on rewards.

probability distribution approximation. We do not choose to show perplexity of PEPDS as it transfers empathy of the generated utterance by PDS using the empathetic transfer model whose PPL score of 8.71 is shown in 3. Further, the **LEN** value of 16.87 indicates that the PEPDS generates longer responses as compared to PDS and ARDM, with lengths of 15.03 and 15.73, respectively, whereas RFI yields the best score of all the four. It could be because repetitive reward in PEPDS penalizes the repetitive tokens, resulting in shorter responses than RFI, but the empathetic transfer in PEPDS results in more extended responses than PDS. **PolSt** score of 41.117% shows that politeness-strategy consistency reward encourages PEPDS to generate the utterances, adapting to the right politeness strategy. Lastly, **Emp** score of 78.1% shows that PEPDS can transfer the empathy in utterances.

Human Evaluation: It can be observed in Table 5 that PEPDS obtains better values for all human evaluation performance measures as compared to the baselines, ARDM and RFI. Scores of **Const**: 4.56, **Fluen**: 4.78, **Adeq**: 3.84 and **N-Rep**:3.78 imply that dialogue-coherence and non-repetitiveness rewards have played a critical role in generating consistent, fluent, adequate and non-repetitive utterances. Further, in terms of **Per**, **Emp** and **Pol-Con**, PEPDS attains the scores of 3.77, 4.21, and 4.03, respectively. Hence, it can be inferred that adding empathy and politeness factor may engage users more in the dialogue. Lastly, it is seen that 67% times users agreed to donate, depicting that our model can effectively connect with the end-user and persuade their donation. More detailed discussions are included in the section A.5 of the appendix.

5.1 Ablation Study

To find the importance of each of the rewards, we experiment with various sets of weights (γ_1, γ_2, γ_3, γ_4 and γ_5) by excluding one reward at a time. These rewards are validated on a 10% held out data. The ablation study to showcase the importance of each weight is shown in Table 6. It can be observed from the table that better perplexity is obtained when all the five rewards are considered, whereas each of its removal affects the perplexity value (i.e. increase in the value). It is also to be noted that persuasion reward γ_1 helps achieve better perplexity when considered with emotion γ_2 and politeness-strategy consistency rewards γ_2. It can be because aspects of politeness and emotion force the agent to generate more engaging utterances with right tone.

Model	Per	Emp	Pol-St	DonPr	Const	Fluen	Adeq	N-Rep
ARDM	2.33	-	-	0.50	3.95	4.17	-	3.17
RFI	2.98	-	-	0.61	4.17	4.41	-	3.50
PDS	3.39	3.91	3.86	0.64	4.48	4.65	3.51	3.72
PEPDS	**3.77**	**4.21**	**4.03**	**0.67**	**4.56**	**4.78**	**3.84**	**3.78**

Table 5: Results of human evaluation

ABLATION STUDY					
γ_1	γ_2	γ_3	γ_4	γ_5	**PPL**
0	0	0	0.5	0.5	11.3100
0	0	0.8	0.1	0.1	11.2830
0	0.8	0	0.1	0.1	11.3123
0.8	0	0	0.1	0.1	11.1164
0.3	**0.3**	**0.2**	**0.1**	**0.1**	**11.0671**

Table 6: Ablation study to showcase the effectiveness of all the five sub-rewards

6 Conclusion

In persuasive conversations, even responses with factual arguments and the right cause may not be able to persuade due to a lack of polite and empathetic tone. Therefore, to ensure these two aspects in persuasive messages, we first fine-tune an MLE loss language model with an RL-loss function consisting of five rewards *viz.* persuasion, emotion, politeness-strategy consistency, dialogue-coherence, and non-repetitiveness. Then generated non-empathetic utterances are transferred using an auto-regressive seq2seq model to empathetic utterances. Detailed empirical evaluation concerning both automatic and human evaluation metrics demonstrate that our proposed model, PEPDS can achieve state-of-the-art performance compared to the existing baselines and can retain both the aspects of emotion and politeness-strategy consistency at par in an ongoing dialogue. Our results also conclude that adding the empathetic transfer model helps the proposed model better facilitate empathy in persuasive responses.

We would like to see more into the personalization aspects in the future to model persuasion, such as likeness, authority, demography, etc.

7 Ethical considerations

In this work, persuasion is modelled using a publicly available dataset. We adhered to the policies of the dataset and have not violated any copyright issues. Dataset which has been used to model polite empathetic persuasion will be made available only with an official agreement with restriction that data would be used only for research works. The dataset is annotated by recruiting three annotators who were paid as per our university norms. We have also got our data annotation process verified by our university review board. Further, persuasion is an intricate process which can be used for personal gain. Therefore, to develop a persuasive conversational AI an ethical intention must be taken into account. In this work, we choose to style persuasive dialogue generation with politeness and empathy for social work of donation to a charity organization utilizing PERSUASIONFORGOOD dataset.

8 Acknowledgement

Kshtij Mishra gratefully acknowledge the assistance of the Indian government's "Prime Ministers Research Fellowship (PMRF) Program." Authors also acknowledge the partial support from the project "Sevak-An Intelligent Indian Language Chatbot", sponsored by Imprint, SERB..

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.

Dario Bertero, Farhad Bin Siddique, Chien-Sheng Wu, Yan Wan, Ricky Ho Yin Chan, and Pascale Fung. 2016. Real-time speech emotion and sentiment recognition for interactive dialogue systems. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, pages 1042–1047.

Gerd Bohner, Markus Ruder, and Hans-Peter Erb. 2002. When expertise backfires: Contrast and assimilation effects in persuasion. *British Journal of Social Psychology*, 41(4):495–519.

Penelope Brown and Stephen C Levinson. 1978. Universals in language usage: Politeness phenomena. In

Questions and politeness: Strategies in social interaction, pages 56–311. Cambridge University Press.

Penelope Brown, Stephen C Levinson, and Stephen C Levinson. 1987. *Politeness: Some universals in language usage*, volume 4. Cambridge university press.

Peter F Brown, Stephen A Della Pietra, Vincent J Della Pietra, Jennifer C Lai, and Robert L Mercer. 1992. An estimate of an upper bound for the entropy of english. *Computational Linguistics*, 18(1):31–40.

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.

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.

Ritam Dutt, Rishabh Joshi, and Carolyn Rose. 2020. Keeping up appearances: Computational modeling of face acts in persuasion oriented discussions. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7473–7485.

Ritam Dutt, Sayan Sinha, Rishabh Joshi, Surya Shekhar Chakraborty, Meredith Riggs, Xinru Yan, Haogang Bao, and Carolyn Rose. 2021. Resper: Computationally modelling resisting strategies in persuasive conversations. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 78–90.

Mauajama Firdaus, Hardik Chauhan, Asif Ekbal, and Pushpak Bhattacharyya. 2020. Emosen: Generating sentiment and emotion controlled responses in a multimodal dialogue system. *IEEE Transactions on Affective Computing*.

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

Hitesh Golchha, Mauajama Firdaus, Asif Ekbal, and Pushpak Bhattacharyya. 2019. Courteously yours: Inducing courteous behavior in customer care responses using reinforced pointer generator network. 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 851–860.

Navita Goyal, Balaji Vasan Srinivasan, N Anandhavelu, and Abhilasha Sancheti. 2021. Multi-style transfer with discriminative feedback on disjoint corpus. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3500–3510.

Junxian He, Xinyi Wang, Graham Neubig, and Taylor Berg-Kirkpatrick. 2020. A probabilistic formulation of unsupervised text style transfer. *arXiv preprint arXiv:2002.03912*.

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

Di Jin, Zhijing Jin, Zhiting Hu, Olga Vechtomova, and Rada Mihalcea. 2022. Deep learning for text style transfer: A survey. *Computational Linguistics*, 48(1):155–205.

Jacob Devlin Ming-Wei Chang Kenton and Lee Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*, pages 4171–4186.

Ahmet Baki Kocaballi, Shlomo Berkovsky, Juan C Quiroz, Liliana Laranjo, Huong Ly Tong, Dana Rezazadegan, Agustina Briatore, and Enrico Coiera. 2019. The personalization of conversational agents in health care: systematic review. *Journal of medical Internet research*, 21(11):e15360.

Kalpesh Krishna, John Wieting, and Mohit Iyyer. 2020. Reformulating unsupervised style transfer as paraphrase generation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 737–762.

Huiyuan Lai, Antonio Toral, and Malvina Nissim. 2021. Thank you bart! rewarding pre-trained models improves formality style transfer. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 484–494.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880.

Jiwei Li, Will Monroe, Alan Ritter, Dan Jurafsky, Michel Galley, and Jianfeng Gao. 2016. Deep reinforcement learning for dialogue generation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1192–1202.

Chin-Yew Lin and Eduard Hovy. 2003. Automatic evaluation of summaries using n-gram co-occurrence statistics. In *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, pages 150–157.

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*.

Eric Malmi, Aliaksei Severyn, and Sascha Rothe. 2020. Unsupervised text style transfer with padded masked language models. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8671–8680.

Pierre-Emmanuel Mazare, Samuel Humeau, Martin Raison, and Antoine Bordes. 2018. Training millions of personalized dialogue agents. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2775–2779.

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

Mohsen Mesgar, Edwin Simpson, and Iryna Gurevych. 2021. Improving factual consistency between a response and persona facts. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 549–562.

Kshitij Mishra, Mauajama Firdaus, and Asif Ekbal. 2022. Please be polite: Towards building a politeness adaptive dialogue system for goal-oriented conversations. *Neurocomputing*, 494:242–254.

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.

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. 2019. Towards empathetic open-domain conversation models: A new benchmark and dataset. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5370–5381.

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. 2021. Refine and imitate: Reducing repetition and inconsistency in persuasion dialogues via reinforcement learning and human demonstration. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 3478–3492.

Weiyan Shi, Xuewei Wang, Yoo Jung Oh, Jingwen Zhang, Saurav Sahay, and Zhou Yu. 2020. 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. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1509–1519.

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.

Haoyu Song, Yan Wang, Kaiyan Zhang, Weinan Zhang, and Ting Liu. 2021. Bob: Bert over bert for training persona-based dialogue models from limited personalized data. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 167–177.

John R Sparks and Charles S Areni. 2002. The effects of sales presentation quality and initial perceptions on persuasion: a multiple role perspective. *Journal of Business Research*, 55(6):517–528.

John R Sparks and Charles S Areni. 2008. Style versus substance: Multiple roles of language power in persuasion. *Journal of Applied Social Psychology*, 38(1):37–60.

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. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5635–5649.

Qingyang Wu, Yichi Zhang, Yu Li, and Zhou Yu. 2021. Alternating recurrent dialog model with large-scale pre-trained language models. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1292–1301.

Jingjing Xu, Xu Sun, Qi Zeng, Xiaodong Zhang, Xuancheng Ren, Houfeng Wang, and Wenjie Li. 2018.

Unpaired sentiment-to-sentiment translation: A cycled reinforcement learning approach. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 979–988.

Zichao Yang, Zhiting Hu, Chris Dyer, Eric P Xing, and Taylor Berg-Kirkpatrick. 2018. Unsupervised text style transfer using language models as discriminators. *Advances in Neural Information Processing Systems*, 31.

A APPENDICES

A.1 Classifiers

To train the classifiers, first contextual representations $c_{<s>}$ are obtained by feeding sampled batches of persuader's utterances to the classifier. These contextual representations $c_{<s>}$ are then passed through a feed forward network to output a vector containing a scalar value for each of the considered classes. Lastly, a softmax is applied on this vector to obtain the resultant vector having probability values for each class. The highest probability value in the resultant vector represents the predicted class for each of the utterances.

A.2 Dataset Annotation Details

EPP4G: First, we annotate the P4G datset with emotions. To achieve this task we used EMPATHETICDIALOGUES (ED) (Rashkin et al., 2019) dataset consisting of approximately 25k conversations grounded in 32 emotions. Due to imbalanced class distribution in EMPATHETICDIALOGUES dataset, we first boil down the 32 labels to 23 labels by merging similar emotion labels. Details of the merged emotion labels can be seen in Table 7 [6]. Then, we train a RoBERTa-large (Liu et al., 2019) based classifier which is used to predict emotion labels for PERSUASIONFORGOOD dataset. These predicted labels are cross-verified manually for a sample of 4000 utterances by three annotators proficient in English communicative skills [7]. They check the trueness of the predicted emotion labels in persuasive context following the EMPATHETICDIALOGUES dataset guidelines to infer their definitions. All three annotators annotate the mis-classified utterances with right emotion labels for each utterance. A reliable multi-rater Kappa (McHugh, 2012) agreement ratio of 72.1% is found between all three annotators annotations.

In order to avoid inevitable face threatening acts, persuader may adopt any one of the different politeness strategies *viz.* positive politeness, negative politeness or off-record (Brown et al., 1987). We ask same three annotators to annotate the approximatey 5300 persuader's utterances with one of these three politeness strategies. Positive politeness strategies are generally used to make the persuadee feel good about themselves or their interests with a implicit known friendly or equal relationship between persuadee and persuader. Negative politeness strategy try to avoid imposition on the persuadee and presume that there is no known relationship between persuadee and persuader. Off-record strategy tries to give some general information or uses indirect language and relies on the persuadee's interpretation for the information that has been conveyed. Annotators were asked to follow these definitions as per guidelines of Brown and Levinson's politeness theory (Brown et al., 1987) to annotate all the 5300 persuader's utterances. Lastly, for each utterance, maximum voted politeness strategy is chosen out of three labels. In these annotations also, a good multi-rater kappa agreement ratio of 78% is found. We name this P4G dataset annotated with right emotion and politeness strategy labels as **EPP4G** dataset.

Emotion_1	Emotion_2	Merged_Emotion
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

Table 7: Emotion classes after merging

ETP4G: To build empathetic transfer model, we required a seq2seq dataset consisting of non-empathetic-to-empathetic utterances. To prepare such a dataset, we follow (Krishna et al., 2020) approach. First, we ask the same three annotators to boil down these 23 emotion labels to only two labels i.e. empathetic and non-empathetic. We ask them to analyse all the 23 emotion labels in EPP4G dataset utterances and discriminate the empathetic and non-empathetic labels to annotate them with respective label. Second, all empathetic utterances are now filtered. It is found that out of 4000 utterances only 817 were non-empathetic and rest 3181 utterances were empathetic. This led to highly imbalanced class distribution [8]. Hence, we oversample each of the non-empathetic utterance using forward-backward machine translation (*English* $\rightarrow$ *Chinese* $\rightarrow$ *English*) with two corresponding similar responses. Hence now, we have $847 \times 3 = 2451$ non-empathetic utterances and 3181 empathetic

[6] The emotion classifier with 32 and 23 labels gives the accuracy scores of 58.17% and 67.44% respectively.

[7] All three annotators were post-graduate qualified and were paid as per our university norms.

[8] A RoBERTa-large (Liu et al., 2019) based empathetic classifier trained on this highly imbalanced dataset yielded the accuracy of 82.5% and macro-f1 of 73.0%

Dataset	#Utterances	Oversampling Details
EMP4G	5632	1. Oversample non-empathetic utterances 2. Use forward-backward machine translation to get more number of utterances 3. #non-empathetic utterances : $817 \times 3 = 2451$ utterances
ETP4G	16722	1. Train a seq2seq paraphrase transformer based model on PARANMT-filtered (Krishna et al., 2020) 2. Filter the empathetic utterances from empathetic P4G dataset. 3. Through forward-backward machine translation increase the size of empathetic utterances: $3181 \times 3 = 9543$. 4. Pass these 9543 empathetic utterances to seq2seq paraphrase transformer based model and get corresponding non-empathetic utterances. 5. Out of 9543 approx utterances, 3287 generated utterance were non-empathetic. 6. Increase the 3287 utterances employing forward-backward machine translation. 7. First, obtain two similar candidates for non-empathetic utterances keeping corresponding empathetic utterances as the same, hence #seq2seq non-empathetic-to-empathteic utterances: $3287 \times 3 = 9861$. Second, keep non-empathetic utterances same and obtain corresponding two similar candidates of empathetic utterances. Hence, #utterances: $3287 \times 5 = 16722$

Table 8: Empathetic transfer model dataset statistics and details

utterances [9]. This dataset is named as **EMP4G** dataset. Third, we train a BART-large seq2seq paraphrase transformer model considering filtered PARANMT dataset with approximately 75k utterances (Krishna et al., 2020). Fourth, in EMP4G dataset the number of empathetic utterances are increased to $3181 \times 3 = 9543$ using same forward-backward machine translation. Now, we pass these increased empathetic utterances to trained seq2seq paraphrase transformer model and obtain the corresponding non-empathetic or empathetic utterances [10]. We considered only these 3,287 non-empathetic utterances and corresponding empathetic utterances to further increase them to $3287 \times 5 = 16722$ by employing forward-backward machine translation.

We call this dataset as **ETP4G** dataset whose details are shown in Table 8. Lastly. these non-empathetic utterances are considered as input and corresponding empathetic utterances as output to train the empathetic transfer model.

[9] A RoBERTa-large (Liu et al., 2019) based empathetic classifier trained on this balanced empathetic dataset gave the accuracy of 85.1% and macro-f1 of 84.6%

[10] Out of 9543 utterances, it was found that 3287 utterance generated were non-empathetic

A.3 Implementation Details

We first used three models: persuasion strategy, emotion and politeness-strategy classifiers to design reward feedback for our RL-agent and then at inference time we use three models: empathetic classifier and empathetic transfer model to generate empathetic utterance.

A.3.1 Classifiers

All four classifiers are trained using transformer based BERT-large: 24-layer, 1024-hidden units, 16-heads and 340M parameters (Kenton and Toutanova, 2019) and Roberta-large: 24-layer, 1024-hidden units, 16-heads and 355M parameters (Liu et al., 2019). For both the architectures, we used batch size = 16 (experimented with 8, 16 also), epochs=2 learning_rate = 4e-5, optimizer = AdamW, attention_dropout = 0.1 and activation = gelu.

A.3.2 Empathetic Transfer Model

We use two seq2seq generation models: BERT-BERT considering encoder and decoder both as BERT (Kenton and Toutanova, 2019) and BART-large having a BERT like encoder and GPT-2 like decoder (Lewis et al., 2020). BART consists of

24 layers, 1024 hidden units, 16 heads and 406M parameters. We followed following configuration to train our BART-based model: optimizer = AdamW, learning_rate = 4e-5, activation = gelu, attention_dropout = 0.1, repetition_penalty = 1.0 and max_seq_length = 128.

A.3.3 Device configurations details

To train transformer based classifiers, empathetic transfer model and PEPDS, we used following device configurations:

- **GPU:** A100-PCIE-40GB.
- **CUDA Support:** CUDA 11.x (or later.
- **Memory clock:** 1215 MHz.
- **Total board power:** 250 W.
- **GPU clocks:** Base: 765 MHz, Boost: 1410 MHz.
- **Memory Size:** 40 GB.
- **Memory Type:** HBM2.
- **Bus Width:** 5120 bits.

A.4 Weight Optimization

In order to find the right combination of weights for the reward function, we investigate with various sets of weights (γ_1, γ_2, γ_3, γ_4 and γ_5). These rewards are validated on a 10% held out data. Finally, the weight combination giving the highest perplexity is chosen. Considered weights for optimization is shown in Table 9. It can be observed from table that better perplexity is obtained when balance between all five rewards *viz.* persuasion, emotion, politeness-strategy consistency, dialogue-coherence and non-repetitiveness is taken into account showcasing the importance of each of the rewards used. It can also be inferred that when emotion and persuasion reward kept constant with keeping a balance between all three rewards, the PPL score varied near 11.07, showcasing the dependency between all three rewards: persuasion, emotion and politeness-strategy consistency. Whereas when any of these three rewards are not taken into account an abrupt drop in PPL score is seen.

REWARD WEIGHT OPTIMIZATION					
γ_1	γ_2	γ_3	γ_4	γ_5	**PPL**
0	0	0	0.5	0.5	11.3100
0	0	0.8	0.1	0.1	11.2830
0	0.8	0	0.1	0.1	11.3123
0.8	0	0	0.1	0.1	11.1164
0.3	**0.3**	**0.2**	**0.1**	**0.1**	**11.0671**
Keeping Emotion constant					
0.1	0.1	0.6	0.1	0.1	11.0785
0.3	0.1	0.4	0.1	0.1	11.0737
0.4	0.1	0.3	0.1	0.1	11.0740
0.5	0.1	0.2	0.1	0.1	11.0738
Keeping Persuasion constant					
0.2	0.3	0.3	0.1	0.1	11.0662
0.2	0.2	0.4	0.1	0.1	11.0697
0.1	0.2	0.5	0.1	0.1	11.0711
0.1	0.4	0.3	0.1	0.1	11.0813

Table 9: Weight Optimisation using different values of γ

A.5 Discussion

An example of user's interaction with proposed PEPDS and ARDM model is shown in Table 10. It can be seen that the user gave better ratings in terms of Per, Emp, Const, Fluen, Adeq, N-Rep to PEPDS than the ARDM. It is also observed that our proposed system is able to employ different persuasion strategies, such as Credibility Appeal, Emotional Appeal, Foot-in-the-Door, Task-related Enquiry and Self-Modelling as well as acknowledging the underlying emotion like caring, trusting, sentimental etc. Further, user gave score to the Pol-St metric as 3 which showcases that our model is able to adapt right politeness strategy as per the context of the dialogue. These results qualitatively shows that PEPDS is able to generate richer, coherent, engaging, right-emotion oriented and politeness-strategy consistent responses. Further, it can also be seen that once the non-empathetic response is generated by PDS, PEPDS is able to generate much more empathetic utterances utilising empathetic transfer model.

Some limitations are also observed such as in reply to utterances like *'That's good, do they take money orders or checks for donations?'*, model replies that *'You can choose any payment you'd like to make. Any amount from $0 to all your payment'*. This can be due to the fact that defined reward function is not able to cover the crucial aspects of knowledge retainment, and grounding of valuable information in utterances. Hence, we may need knowledge grounding to generate better utterances. Lastly, an another user interaction example with PEPDS examples is given in Table 11.

Speaker	PEPDS model generated utterances (Per=5, Emp=3, Pol-St=3, Const=5, Fluen=5, Adeq=4 and N-rep=4)
PEPDS	Hi. How are you doing today?
USER	I am doing well! How about you?
PEPDS	I am doing pretty well. thanks for asking! Let's talk about donation to a charity organization.
USER	Alright. What you want to talk about?
PEPDS	Good morning. Would you like to donate to a charity called Save the Children?
USER	I have heard of that group but I have never really thought about donating to it.
	Have you heard about a charity organization Save the Children?
USER	Only in name. I haven't heard too much about it though. Can you tell me more?
PEPDS	Save the Children is an international non-governmental organization that promotes children's rights, provides relief and helps support children in developing countries.
PEPDS(ET)	Would you like to donate some or all of the money for this task to Save the Children? Save the children is an international non-governmental organization that promotes children's rights, provides relief and helps support children in developing countries.
USER	That sounds like a good charity. I prefer to keep things closer to home though. In which countries do they operate. Are they active in India too?
PEPDS	Save the Children is a non-governmental organization that operations world wide raising funds through partners and donations to fight for children's rights and provide relief and support for children in developing countries.
PEPDS(ET)	Sure, I can tell you about that. Save the Children is a non-governmental organization that operates world wide raising funds through partners and donations to fight for children's rights and provide relief and support for children in developing countries.
USER	It sounds good. Stiil, I've to think about it.
PEPDS	Have you ever donated to any charities before?
PEPDS(ET)	That is understandable. Have you ever donated to any charity before?
USER	I have, I have been one for about 2 years now. I come on every once and a while and do it part time. I like to call it picking up nickels and dimes. How about yourself?
PEPDS	I have. I usually donate at least $2 every month. Do you feel the same way?
USER	Yes, monthly or yearly I try to donate. Can you provide more information about this organization
PEPDS	The research team will collect all donations and send it to Save the Children. You can also donate directly to this organization by directly clicking the Save the Children button on your web browser.
PEPDS(ET)	The research team will collect all donations and send them to Save the Children. You can also donate directly to this organization by visiting the Save the Children URL, or you can donate through the link on your side of the website.
USER	That's good, do they take money orders or checks for donations?
PEPDS	You can choose any payment you'd like to make. Any amount from $0 to all your payment.
USER	Ok, well I think I could donate half my earnings to this.
PEPDS	That's a good idea. Thank you for your time and willingness to donate.
USER	I think I'll donate $1.
PEPDS	Thank you so much. Have a nice day.
USER	Good to talk to you.
Speaker	**ARDM model generated utterance (Per=2, Emp=2, Pol-St=2, Const=4, Fluen=4, Adeq=2 and N-rep=2)**
ARDM	Hello
USER	Hello how are you doing?
ARDM	Good, have you heard of the Save the Children charity?
USER	No, I have not heard about them.
ARDM	Are you familiar with it?
USER	No I don't know. Please tell me.
ARDM	I am just asking you about Save the Children charity because the children are dying and the government is helping them.
USER	Yes, I understand. That is very sad to hear. What we can do?
ARDM	Do you know about children's charities?
USER	No, I do not know much about them.
ARDM	Do you like to help children?
USER	Definitely, I do like to help them.
ARDM	Do you know children's charities?
USER	I have very little knowledge about them.
ARDM	Do you know about save the children charity?
USER	Like i said very less.
ARDM	are you familiar with the charity?
USER	Thank you very much, bye
ARDM	I will ask you again later
USER	thank you.

Table 10: An example of user interaction with our proposed PEPDS and ARDM. Yellow highlighted responses show the responses with persuasion strategy.

Speaker	PEPDS model generated utterances (Per=5, Emp=4, Pol-St=3, Const=5, Fluen=5, Adeq=3 and N-rep=4)
PEPDS	hi
PEPDS(ET)	Hello, How are You?
USER	Good Morning, how is it going?
PEPDS	Very well thank you for asking
USER	Good, What we are going to talk about?
PEPDS	We are going to talk about a charity called Save the Children.
PEPDS(ET)	We will talk about a charity called Save the Children. Have you heard of this charity before?
USER	Yes, I may have probably not heard about them.
PEPDS	they help children in developing countries like Syria and Uganda.
PEPDS(ET)	They help children mainly in developing countries like Syria and Uganda.
USER	Can you tell me more about the charity?
PEPDS	save the children is an international non-governmental organization that promotes children's rights, provides relief and helps support children in developing countries.
PEPDS(ET)	Sure, Save the Children is an international non-governmental organization that promotes children's rights, provides relief and helps support children in developing countries.
USER	Alright, Good, What would you like from me in this regard?
PEPDS	Well, I am going to ask you something. Would you like to help them out by donating some of your payment for this task to them?
USER	Why you want me to donate for this task?
PEPDS	Well, you know, there is a lot of poverty in the developing world.
USER	Yes I think so, we are so involved in ourselves. What you suggest, What we can do?
PEPDS	We can donate. This organization has done quite a lot for the children in war torn areas. They help with education, health, safety, etc
PEPDS(ET)	We can donate for the children's in need. They help children in war torn areas with education, health care, safety, etc.
USER	Sure. What can we do to help these children?
PEPDS	Okay, Well, we could donate a portion of the payment you are going to get from this task to them.
PEPDS(ET)	Okay, well if you decide you would like to donate any of the proceeds from this task to them, we can donate a portion of the payment you will get from that portion.
USER	I think it is very important to help people in need whenever one can but is this organization legit?
PEPDS	Yes, this organization is completely legit. They have been around for years and I have been working with them for the past two years now.
USER	Oh they are doing god's work. Which countries do they operate in?
PEPDS	I think most of them are in Africa. They are actually based out of Nairobi, Kenya. I have been working with them for the past two years now and I have not heard anything but good things from them.
USER	I am not sure if I will donate. I need more information about the charity.
PEPDS	You are absolutely right to doubt. But, It is a great charity and it has helped save countless lives. It is a great cause, you can donate any amount from $0 to all your payment for this task.
USER	Can you tell me more about the charity?
PEPDS	It is an international non-governmental organization that promotes children's rights, provides relief and helps support children in developing countries.
PEPDS(ET)	It would be pleasure to tell you about it! it is an international non-governmental organization that promotes children's rights, provides relief and helps support children in developing countries.
USER	Can you tell me how the money will be collected and where it will go?
PEPDS	The research team will collect all donations and send it to Save the Children. Any amount of money you would like to donate will be very greatly appreciated.
USER	I'm not sure if should i donate?
PEPDS	Do you really care about children?.
USER	Yes, I do care a lot about them and love them.
PEPDS	It is a great charity and every little bit helps. Your donation will help children in need and in this task, you can help them as well.
PEPDS(ET)	That's perfect. It is a great charity and every little bit helps. Your donation will help children in need and in this task, you can help them as well.
USER	Sure, I'll donate. I can think of donating some of my earnings with this task.
PEPDS	That's great, I appreciate your donation. Thank you again for your help.
USER	Yes, we should help as much as possible.
PEPDS	How much amount do you like to donate to the charity?
PEPDS(ET)	Yes, Great. How much would you like to donate to charity now?
USER	0.5$
PEPDS	Okay that's great. Thanks again for your time. Have a wonderful day.

Table 11: An example of user interaction with our proposed PEPDS. Yellow highlighted responses show the responses with persuasion strategy.