

Understanding User Resistance Strategies in Persuasive Conversations

Youzhi Tian¹, Weiyan Shi¹, Chen Li², Zhou Yu¹

¹University of California, Davis, ²Tencent
{yotian, wyshi, joyu}@ucdavis.edu
ailabchenli@tencent.com

Abstract

Persuasive dialog systems have various usages, such as donation persuasion and physical exercise persuasion. Previous persuasive dialog systems research mostly focused on analyzing the persuader’s strategies, and paid little attention to the persuadee (user). However, understanding and addressing users’ resistance strategies is an essential job of a persuasive dialog system. So, we adopt a preliminary framework on persuasion resistance in psychology, and design a fine-grained resistance strategy annotation scheme. We annotate the PERSUASIONFORGOOD dataset with the scheme (Wang et al., 2019). With the enriched annotations, we build a classifier to predict the resistance strategies. Furthermore, we analyze the relationships between persuasion strategies and persuasion resistance strategies. Our work lays the ground for developing a persuasive dialogue system that can understand and address user resistance strategy appropriately. The code and data will be released.

1 Introduction and Related Work

Persuasion plays a prominent role in human communication and has attracted more and more attentions in the NLP community (Tan et al., 2016; Hidey et al., 2017; Hidey and McKeown, 2018; Wang et al., 2019; Yang et al., 2019). During persuasion, the persuader attempts to convince the persuadee to change his/her attitude, opinion or behavior. Previous research on persuasive dialogs mainly study the persuader’s strategies (Wang et al., 2019; Shi et al., 2020; Li et al., 2019), which is helpful when the persuadee shows positive attitude towards the persuasion; but in other situations, the persuadees resists rather than embrace the persuasive attempt. For instance, in PERSUASIONFORGOOD (Wang et al., 2019), 166 out of 1,017 persuasive dialogs contains resistance strategy; although individuals resist persuasive attempts with different tac-

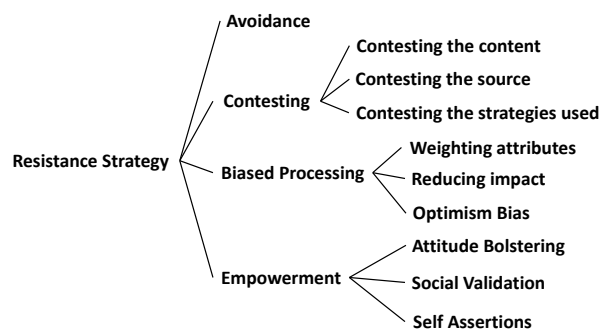


Figure 1: The preliminary resistance strategy framework Fransen et al. (2015)

tics, these resistant utterances all collapse into one dialog-act “negative-reactions-towards-persuasion” in the provided annotation scheme, which makes it harder for the persuasive system to respond accordingly. Therefore, to achieve more efficient persuasion, we propose to study the resistance to persuasion in more details.

Persuasion resistance has been studied in social psychology, marketing and so on (Knowles and Linn, 2004; Dal Cin et al., 2004; Petty and Cacioppo, 1977; Ahluwalia, 2000; Haugtvedt and Petty, 1992). Rucker et al. (2004) showed that there exist individual differences in persuasion resistance dependent on their beliefs. Further, one preliminary framework in psychology (Fransen et al., 2015) categorized the resistance behaviors into four main categories: *Avoidance*, *Contesting*, *Biased processing*, and *Empowerment*, as shown in Figure 1. In our work, we adopt this framework as a guidance, and design a more fine-grained annotation scheme with resistance strategies for PERSUASIONFORGOOD. Two annotators label 447 persuadee resistant utterances in the dataset with six different resistance strategies. With these enriched labels, we build a classifier to predict the resistance strategies for further persuasive dialog system development.

To perform more efficient persuasion with the

Category	Amount
Avoidance	127
Self assertion	114
Weighting attributes	116
Contesting content	84
Attitude bolstering	5
Contesting strategy	1
Total	447

Table 1: Statistics of persuasion resistance strategies in PERSUASIONFORGOOD.

resistance strategy, we also need to study the relationship between persuader’s persuasion strategy and persuadee’s resistance strategy. Therefore, we analyzed such relationship in PERSUASIONFORGOOD, which suggests that while *Self-assertion* is generally difficult to handle, other resistance strategies such as *Weighting attributes* sometimes can be addressed by persuasion strategies such as *credibility appeal*. These insights will play an important role in the future design of a persuasive dialog systems with user resistance strategy awareness.

2 Resistance Strategy Annotation Scheme and Analysis

Based on the framework in psychology (Fransen et al., 2015), we design a fine-grained resistance strategy annotation scheme. As this persuasion task was complicated, different utterances in the same dialog turn may have different semantic meanings. We decide to annotate each utterance instead of the dialog turn, following Wang et al. (2019).

We identified six out of ten resistance strategies in the preliminary framework (Fransen et al., 2015) occurred in the PERSUASIONFORGOOD (with the exception of *Contesting the source*, *Reducing impact*, *Optimism Bias* and *Social Validation*), and designed an annotation scheme accordingly. We hired two experts with persuasion theory training to annotate the PERSUASIONFORGOOD based on the scheme. They reached an inter-code reliability of Kappa value of 0.806 for PERSUASIONFORGOOD. We shows the details of annotation scheme below together with some analyses.

Avoidance: Avoidance refers to avoiding persuasion attempts. This is one of the most common and straightforward strategies to protect oneself from the impact of persuasive messages. Table 1 shows that it happens the most in our dataset.

e.g:Persuader: Do you have any more questions for me? (1) Persuadee: Nope. (2) Persuadee: Let’s leave it at that. In the first example, the persuadee uses avoidance strategy by showing no interests on the topic. In the second example, the persuadee directly avoids chatting with persuader on this topic.

Self-assertion: People may resist persuasion by asserting their attitude. When they apply this strategy, they remind themselves that nothing can change their attitudes or behavior because they are confident about them. (E.g., Persuader: Any chance you would consider making a greater donation? (1) Persuadee: This is all I can afford (2) Persuadee: I can’t afford much.(3) Persuadee: Unfortunately , I can’t donate anything right now.(4) Persuadee: 0.60 still sounds good to me.) All these four examples show that the persuadees assert themselves that they would not change their attitudes. When they apply self-assertion strategy, it is difficult for the persuader to persuade them.

Weighting-attributes: Weighting attributes means more weight is attached to information that is consistent with one’s attitudes and less weight is attached to inconsistent information. (E.g: (1)Persuader: It’s a fact that small donations by large numbers of people can have an impact. Persuadee: how do we know the donation aren’t simply buying another yacht for an executive? (2) Persuader: It doesn’t take much to save 30c a week. Persuadee: that is true, but its hard to justify that 30c is going to make any difference.) Both examples show that persuadee attach more weights to the side effects that may occur after they donate the money.

Contesting-content: Contesting content means people reflect on the augments in the message and subsequently use counterarguments to refute it. It is a thought process that decreases agreement with a counter attitudinal message. (Eg: (1) Persuder: Would you consider to donate? Persuadee: I don’t think so. (2) Persuader: they are a good group. Persuadee: That doesn’t sound very good.) Both examples show that persuadees refute the content from persuaders.

Attitude-bolstering: Attitude bolstering is a process by which people generate thoughts that are supportive of their existing attitudes and behavior. (Eg: Persuadee: As they say, don’t light yourself on fire to keep others warm.) This example shows that the persuadee uses the existing attitude to refuse the persuasion. Table 1 shows that the attitude bolstering only occurs five times in

	Avoidance	Self-assertion	Weighting-attributes	Contesting content
Avoidance	0.75	0.05	0.10	0.10
Self assertion	0.11	0.67	0.09	0.12
Weighting attributes	0.28	0.09	0.49	0.14
Contesting content	0.09	0.07	0.09	0.74

Table 2: Confusion matrix for the four resistance strategies using the RCNN classifier.

Category	Donation Rate
Avoidance	21% (20/93)
Self assertion	15% (13/82)
Weighting attributes	26% (21/79)
Contesting content	23% (16/68)
Attitude bolstering	25% (1/4)
Contesting strategy	0% (0/1)
all dialogs	54% (545/1017)
dialogs with strategies	22% (37/166)

Table 3: Statistics of donation rate of dialogs with different resistance strategies in PERSUASIONFORGOOD.

PERSUASIONFORGOOD . It is mainly because this resistance strategy is relatively indirect compared with other resistance strategies.

Contesting-strategy: contesting strategy means people develop theories and beliefs about how persuasion agents try to influence them. (Eg: Persuader: So one last time how much exactly would you like to donate to the charity now? Persuadee: Pushiness does not make for a good salesperson.) In this example, the persuadee has developed their belief that the persuader tries to push for donation. Then the persuadee contests the strategy that the persuader used. Table 1 shows that the contesting strategy only occurs once. It is not surprising that most of the persuadees have no experience on the persuasion. They have not developed beliefs about how persuaders try to persuade them.

3 How Resistance Strategies Influence Donation Rate

Table 1 shows the statistics of resistance persuasion strategies in PERSUASIONFORGOOD dataset. There are altogether 166 out of 1,017 persuasive dialogs containing resistance strategy. With the enriched dataset, we first want to learn whether the resistance strategy can affect the persuasion result. So we calculate the donation rate for the dialogs

with each resistance strategy.

Table 3 shows that the donation rate in the whole dataset is 54% (545 out of 1017), while the donation rate in the dialogs which contain resistance strategy is only 22% (37 out of 166). The low donation rate shows that with the resistance strategy, it is more difficult for the persuader to persuade the users, indicating that the resistance strategy has an important effect to the persuasion result. Table 3 also shows the donation rate of the dialogs with certain resistance strategy. As some dialogs may contain several resistance strategies, we calculate the donation rate in different categories separately. We find that dialogs with *Self assertion* strategy have the lowest donation rate. It is mainly because persuadees who apply this strategy remind themselves that nothing can change their attitudes. They reinforced their not-donating attitude confidence by using the strategy. For example, the persuadee said “Unfortunately I can’t donate anything right now.” which shows strong confidence. Then the persuader said “it is really sad to hear.” Such resistant strategies are difficult for persuaders to handle.

The dialogs with *weighting attributes* strategy have the highest donation rate. It is not surprising as the persuaders can use multiple persuasion strategies such as credibility appeal and logical appeal to eliminate the persuadees’ concerns about certain attributes. For example, when persuadee said “how do we know the donation aren’t simply buying another yacht for an executive? ”, he has the concern for the usage of the donation. Then the persuader said “A review of the tax exempt paperwork and financial statements can help you weed out good from bad.” The persuader here uses credibility appeal strategy to address the persuadee’s concern. So if the persuasive dialog system can classify different resistance strategies and apply different persuasion strategies to deal with certain resistance strategies, it can effectively increase the persuasiveness of the conversational agent.

4 Resistance Strategy Classification

In order to build a persuasive dialog system with user resistance strategy awareness, we need to first understand the resistance persuasion patterns and differentiate various resistance persuasion strategies from each other. So we designed a classifier for the resistance persuasion strategies in PERSUASIONFORGOOD dataset.

4.1 Model

Table 1 shows that the utterances mainly fall into six categories: *Avoidance*, *Self assertion*, *Weighting attributes* and *Contesting content*. But the number of times each strategy appears is unbalanced especially for *Attitude bolstering* and *Contesting strategy*. The imbalanced distribution in the dataset increases the difficulty in training a resistance strategy classifier. As the *Attitude bolstering* is similar to *Weighting attributes* and *Contesting strategy* is similar to *Contesting content*, we combine these categories together to balance the data distribution. So there are in total four categories with relatively balanced samples. Then we can train a classifier to predict these resistance strategies.

We use recurrent convolutional neural network (RCNN) to train the resistance strategy classifier. Recurrent convolutional neural network (RCNN) combine CNN and RNN to extract both the global and local semantics, and the recurrent structure may reduce noise compared to the window-based neural network (Lai et al., 2015). We use the hidden state of the LSTM with a hidden size of 200 as the sentence embedding. Then a linear semantic transformation is applied on the sentence embedding to obtain the input to a max-pooling layer. Finally the pooling layer is used to capture the useful information throughout the entire sentence.

There are 447 utterances with resistance persuasion strategy. We split the utterances into 357 for training, 45 for validation and 45 for test. We train the classifier for 20 epochs and choose the one that performs best on the validation dataset. Adam is used for the optimization. We set the initial learning rate to be 0.001 and applied exponential decay every 100 steps. In addition, dropout (Srivastava et al., 2014) with a probability of 0.5 was applied to reduce over-fitting.

4.2 Experiments and Analysis

Experiments show our classifier’s accuracy is 67.34% and F1 score is 0.52. Since our dataset

is unbalanced in categories, so in addition to accuracy, we also use the macro F1 as the evaluation metric. The F1 score is 0.522. The experimental results suggest that we can classify different resistant strategies with good performance. We plan to incorporate the resistance persuasion strategy classifier into persuasive dialog systems in the future.

To further analyze the results, we plotted the confusion matrix for the four resistance strategies in Table 2. We found the main error comes from the misclassification of *Weighting attributes*. Sometimes sentences of *Weighting attributes* were misclassified as *Avoidance*, because when the persuadees use *Weighting attributes*, they may use avoidant words to avoid the attributes which are inconsistent with their attitudes. For example, in the sentence “Let’s leave it because I do not know where my donation will go.” the persuadee attach more weights to the concern that the donation may be used inappropriately. The persuadee also uses avoidant words “leave it” to avoid the donation request that persuader proposed. In summary, one sentence may convey multiple meanings which led to misclassifications due to the complex nature of human-human persuasion dialogs.

5 Conclusions and Future Work

In this paper, we introduce a persuasion resistance strategy annotation scheme to assist persuasion dialog analyses and persuasive dialog system building. We annotate the PERSUASIONFORGOOD dataset with the resistance strategy annotation scheme. We find that users who use resistance strategies are less likely to donate compared with users who don’t show resistant attitudes. Further, different resistance strategies have different impacts on the donation outcome. For example, it is more difficult for the persuader to handle with the *Self assertion* strategy than the *Weighting attributes* strategy. We also built a baseline classifier to classify different resistance persuasion strategies. Results show that the classifier can obtain relative high accuracy and F1 score. Such classifier can assist persuasive dialog systems to identify different resistance strategies and choose corresponding persuasion strategies to deal with them. It is also helpful to build a user simulator when testing the persuasive dialog system. In conclusion, this study marks the first step to incorporate resistance strategy into a persuasive dialog system for better persuasion outcomes.

References

- Rohini Ahluwalia. 2000. Examination of psychological processes underlying resistance to persuasion. *Journal of Consumer Research*, 27(2):217–232.
- Sonya Dal Cin, Mark P Zanna, and Geoffrey T Fong. 2004. Narrative persuasion and overcoming resistance. *Resistance and persuasion*, 2:175–191.
- Marieke L. Fransen, Edith G. Smit, and Peeter W. J. Verlegh. 2015. [Strategies and motives for resistance to persuasion: an integrative framework](#). *Frontiers in Psychology*, 6:1201.
- Curtis P Haugtvedt and Richard E Petty. 1992. Personality and persuasion: Need for cognition moderates the persistence and resistance of attitude changes. *Journal of Personality and Social psychology*, 63(2):308.
- Christopher Hidey, Elena Musi, Alyssa Hwang, Smaranda Muresan, and Kathleen McKeown. 2017. Analyzing the semantic types of claims and premises in an online persuasive forum. In *Proceedings of the 4th Workshop on Argument Mining*, pages 11–21.
- Christopher Thomas Hidey and Kathleen McKeown. 2018. Persuasive influence detection: The role of argument sequencing. In *Thirty-Second AAAI Conference on Artificial Intelligence*.
- Eric S Knowles and Jay A Linn. 2004. *Resistance and persuasion*. Psychology Press.
- Siwei Lai, Liheng Xu, Kang Liu, and Jun Zhao. 2015. Recurrent convolutional neural networks for text classification. In *AAAI*, pages 2267–2273. AAAI Press.
- Yu Li, Kun Qian, Weiyang Shi, and Zhou Yu. 2019. End-to-end trainable non-collaborative dialog system. *arXiv preprint arXiv:1911.10742*.
- Richard E Petty and John T Cacioppo. 1977. Forewarning, cognitive responding, and resistance to persuasion. *Journal of Personality and social Psychology*, 35(9):645.
- Derek D Rucker, Zakary L Tormala, and Richard E Petty. 2004. Individual differences in resistance to persuasion: The role of beliefs and meta-beliefs. *Resistance and persuasion*, page 83.
- Weiyang Shi, Xuewei Wang, Yoo Jung Oh, Jingwen Zhang, Saurav Sahay, and Zhou Yu. 2020. Effects of persuasive dialogues: Testing bot identities and inquiry strategies. *arXiv preprint arXiv:2001.04564*.
- Nitish Srivastava, Geoffrey E. Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15(1):1929–1958.
- Chenhao Tan, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2016. Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In *Proceedings of the 25th international conference on world wide web*, pages 613–624.
- Xuewei Wang, Weiyang 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 *ACL (1)*, pages 5635–5649. Association for Computational Linguistics.
- 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.