Examining the Ordering of Rhetorical Strategies in Persuasive Requests

Omar Shaikh, Jiaao Chen, Jon Saad-Falcon, Duen Horng (Polo) Chau, Diyi Yang
Georgia Institute of Technology
{oshaikh, jchen896, jonsaadfalcon, polo, dyang888}@gatech.edu

Abstract

Interpreting how persuasive language influences audiences has implications across many domains like advertising, argumentation, and propaganda. Persuasion relies on more than a message's content. Arranging the order of the message itself (i.e., ordering specific rhetorical strategies) also plays an important role. To examine how strategy orderings contribute to persuasiveness, we first utilize a Variational Autoencoder model to disentangle content and rhetorical strategies in textual requests from a large-scale loan request corpus. We then visualize interplay between content and strategy through an attentional LSTM that predicts the success of textual requests. We find that specific (orderings of) strategies interact uniquely with a request's content to impact success rate, and thus the persuasiveness of a request.

1 Introduction

Persuasion has been shown as a powerful tool for catalyzing beneficial social and political changes (Hovland et al., 1953) or enforcing propaganda as a tool of warfare (Finch, 2000). Modeling persuasiveness of text has received much recent attention in the language community (Althoff et al., 2014; Tan et al., 2016; Habernal and Gurevych, 2017; Yang et al., 2019; Srinivasan et al., 2019). Numerous qualitative studies have been conducted to understand persuasion, from explorations of rhetoric in presidential campaigns (Bartels, 2006; Popkin and Popkin, 1994) to the impact of a communicator's likability on persuasiveness (Chaiken, 1980). Studies of persuasion and argumentation that have analyzed textual level features (e.g., n-grams, independent rhetorical strategies) to gauge efficacy have also garnered recent attention (Althoff et al., 2014; Habernal and Gurevych, 2017, 2016b,a; Yang and Kraut, 2017; Yang et al., 2019). Of particular interest is Morio et al. (2019), which identified sentence

Strategy	Definition
Concreteness (39%)	Use concrete details in request *"I need $250 to purchase fishing rods"*
Reciprocity (18%)	Assure user will repay giver *"I will pay 5% interest to you"*
Impact (12%)	Highlight the impact of a request *"This loan will help teach students"*
Credibility (8%)	Use credentials to establish trust *"I have repaid all of my prior loans"*
Politeness (16%)	Use polite language *"Highly appreciated."*
Other (7%)	None of the above

Table 1: Sentence level persuasion strategies, and their data distributions (%). Strategy abbreviations are **bolded**.

placements for individual rhetorical strategies in a request. Other research analyzed how different persuasive strategies are more effective on specific stances and personal backgrounds (Durmus and Cardie, 2018, 2019).

However, prior work has mainly focused on identifying overall persuasiveness of textual content or analyzing components of persuasion affecting a request. These works largely ignore **ordering** of specific strategies, a key canon of rhetoric that has a large impact on persuasion effectiveness (Borchers and Hundley, 2018; Cicero, 1862). In the context of online communities, identifying where/how effective orderings occur may highlight qualities of persuasive requests and help users improve their rhetorical appeal. Furthermore, highlighting ineffective orderings may help users avoid pitfalls when framing their posts.

To fill this gap, we propose to investigate particular orderings of persuasive strategies that affect a request's persuasiveness and identify situations where these orderings are optimal. Specifically, we take a closer look at strategies (Table 1) and their orderings in requests from the subred-
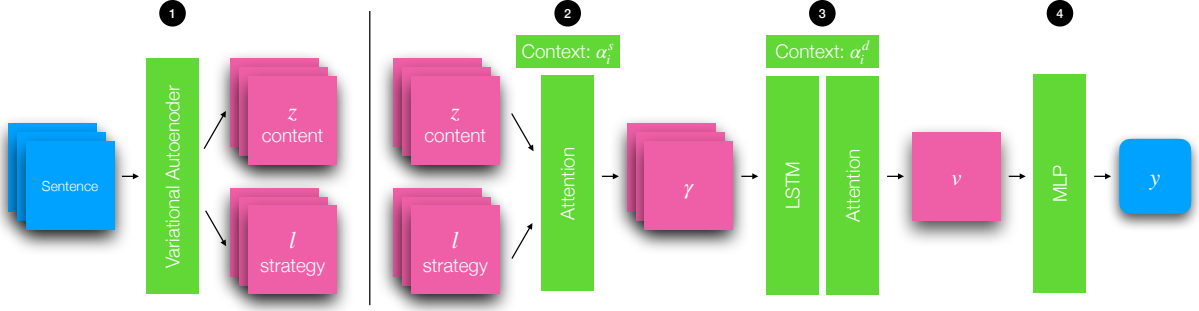
Figure 1: Our modeling setup, detailed in section 2, consists of 4 steps. Step 1 deconstructs sentences into latent content and strategy vectors, using a semi-supervised VAE; Step 2 combines content and strategy vectors at the sentence level, using sentence level attention; Step 3 uses an LSTM to model our sentences in a request, then combines sentences using request level attention. Finally, Step 4 predicts our binary persuasiveness label using a multilayer perceptron.

dit/online lending community **r/Borrow** [1]; and utilize them to examine research questions like: When should requesters follow strategy orderings (e.g., ending loan requests with politeness) that rely on social norms? Should requesters worry less about orderings and more about content? Altogether, this work examines **orderings**, an overlooked rhetorical canon, and how they interact with a request's persuasiveness in an online lending domain. Our contributions include:

1. Identifying specific strategy orderings that correlate with requests' persuasiveness.

2. Highlighting the interplay between content and strategy with respect to the persuasiveness of a request.

3. Perturbing underperforming strategy orderings to help improve persuasiveness of requests via a set of introduced edit operations.

Code for our analyses can be found at https://github.com/GT-SALT/Persuasive-Orderings.

2 Method

2.1 Dataset

Our Borrow dataset consists of 49,855 different loan requests in English, scraped from the r/Borrow subreddit. r/Borrow is a community which financially assists users with small short-term loans to larger long-term ones. Every request has a binary label indicating if a loan is successful or not. Request success rate, on average, is 48.5%. We randomly sampled a subset (5%) from the whole corpus to annotate their sentence-level labels indicating persuasive strategies; labels were adapted from

Yang et al. (2019) (and defined in Table 1). We recruited four research assistants to label persuasion strategies for each sentence. Definitions and examples of different persuasion strategies were provided, together with a training session where we asked annotators to annotate a number of example sentences and walked them through any disagreed annotations. To assess the reliability of the annotated labels, we then asked them to annotate a small subset of 100 requests from our corpus, with a Cohen' Kappa of .623, indicating moderate annotation agreement (McHugh, 2012). Annotators then annotated the rest of corpus by themselves independently. In total, we gathered 900 requests with sentence-level labels and 48,155 requests without sentence-level labels as our training set, 400 requests with sentence-level labels as the validation set and 400 requests with sentence-level labels as the test set.

2.2 Modeling

Persuasive sentences are combinations of content (what to include in persuasive text) and strategy (how to be persuasive). To explore interplay between content and strategy orderings inside requests, we followed Kingma and Welling (2014) and Yang et al. (2017), utilizing a semi-supervised Variational Autoencoder (VAE) trained on both labeled and unlabeled sentences to disentangle sentences into strategy and content representations. Specifically, for every input sentence x, we assumed the graphical model $p(x, z, l) = p(x|z, l)p(z)p(l)$, where z is a latent "content" variable and l is the persuasive strategy label. The semi-supervised VAE fits an inference network $q(z|x, l)$ to infer latent variable z, a generative network $p(x|l, z)$ to reconstruct input sentence s, and

a discriminative network $q(l|x)$ to predict persuasive strategy l, while optimizing an evidence lower bound (ELBO) similar that of general VAE. We report a Macro F-1 score of 0.75 on the test set for *sentence-level classification*, suggesting reasonable performance compared to an LSTM baseline with a Macro F-1 of 0.74 (Yang et al., 2019). Then, for each request $M = \{x_0, x_1, ..., x_L\}$ consisting of L sentences that a user posted to receive a loan, we utilized our trained semi-VAE to represent each sentence x_i in M with content and strategy variables to form $M' = \{(z_0, l_0), (z_1, l_1)\dots(z_L, l_L)\}$.

With the intent of interpreting importance between strategy orderings and content, we built an attentional LSTM trained to predict success of a request. For each disentangled sentence (z_i, l_i) in our requests, we first applied attention on z_i and l_i at the **sentence level**, dynamically combining them into sentence representation γ_i:

$$u_{z_i} = \tanh\left(W_z z_i + b\right)$$
$$u_{y_i} = \tanh\left(W_l y_i + b\right)$$
$$\alpha_i^s = \frac{\langle \exp\left(u_{z_i}^\top u_q\right), \exp\left(u_{l_i}^\top u_q\right)\rangle}{\exp\left(u_{z_i}^\top u_q\right) + \exp\left(u_{l_i}^\top u_q\right)}$$
$$\gamma_i = \left(\alpha_{i,0}^s \cdot z_i\right) \oplus \left(\alpha_{i,1}^s \cdot l_i\right)$$

where u are randomly initialized context vectors that were jointly learned with weights W. We computed the request representation v through an LSTM that encoded sentence representations γ_i for each request, and a **request level** attention that aggregated information from different sentences. Overall persuasiveness is predicted as:

$$u_i = \tanh\left(W_s h_i + b_s\right)$$
$$\alpha_i^d = \frac{\exp\left(u_i^\top u_s\right)}{\sum_k \exp\left(u_k^\top u_s\right)}$$
$$v = \sum_i \alpha_i^d h_i \text{ and } y = \text{MLP}(v)$$

The training objective is regular cross entropy loss. Macro-averaged performances for *request-level classification* on several baseline classifiers are shown in Table 2. Our attentional model (VAE + LSTM) achieves comparable performance to BERT, while providing additional benefit of disentangling content and strategy. This helps yield relative measures of importance for content and strategies.

3 Interplay of Ordering and Content

To examine how different strategy orderings contribute to overall persuasiveness of requests, we

Model	F-1	Precision	Recall
Naive Bayes	.60	.60	.60
BERT	.65	.64	.65
VAE + LSTM	.61	.61	.61

Table 2: Request label performance on test set.

identified relationships between strategy orderings and success rate by analyzing learned attention weights between strategy orderings and content in our model. Motivated by the "Rule of Three" prevalent in persuasive writing (Clark, 2016), we utilized triplets as our strategy unit of analysis. The most important strategy triplet in each request was considered to be its **"persuasion strategy triplet."**

Pinpointing strategy triplets involved finding the most important consecutive three sentences $((z_{m-1}, l_{m-1}), (z_m, l_m), (z_{m+1}, l_{m+1}))$ in one request based on highest **request-level** (α^d) attention weight associated with a sentence. The strategies (l_{m-1}, l_m, l_{m+1}) associated with these sentences were defined as the aforementioned strategy triples. We noted that the cumulative request-level (α^d) attention placed on strategy triplets had $\mu = .98$ and $\sigma = .07$, indicating that a single triplet carried most responsibility for persuasiveness of requests. For our analysis, we also defined **success rate** of a strategy triplet as the average success rate of the requests it belongs to, irrespective of how important it is to a request (ignoring α^d). To control for infrequent triplets, we defined rare strategies as consisting of less than 0.5% of our dataset. We filtered these rare strategies, along with triplets containing the undefined "Other" strategy. Finally, we averaged **sentence-level** attention weights α^s on each strategy representation in a triplet to represent the importance of an ordering pattern compared to content. Figure 2 plots **sentence-level** attention weights for each strategy triplet and its corresponding success rate.

We made three discoveries. (1) Success rate and triplet attention were *strongly negatively correlated* ($R = -.90$, $p < .0001$). Therefore, the model paying larger attention to strategy triplets may communicate a request's lack of persuasiveness. (2) Attention from around strategy (SOS, Im, Re) onward decreased substantially, suggesting that *content (complementary to strategy attention) played an increasingly larger role in determining the persuasiveness for strategies triplets above average success rate.* (3) Under-performing strategies

Strategy	Example	Strategy Attention	Success Rate
(Po, Po, EOS)	... Your help is **deeply appreciated! Thank you** for reading and considering my request!	.00	.82
(Re, Po, EOS)	... I would greatly appreciate the help, will **pay back latest on 10/31/18 with interest** determined by the lender. **Thank you in advance!**	.01	.76
(Co, Po, EOS)	... I'm trying to budget myself to where I can borrow **less this pay period** and get myself to where I won't need to borrow anymore. Until then I would really **appreciate the help!**	.02	.71
(Co, Im, Co)	... I am switching jobs and need to do **2 weeks off site training.** The store I am training at is several hours away so I need gas money and money **for food** while I am out of town. Everything I spend will be **reimbursed at the end of training**...	.09	.34
(Co, Co, Co)	... Was supposed to be on a flight back to TX **earlier today.** It got canceled **due to weather and delays** (can provide proof). I'm now stuck in Colorado for **another day** before I can get back to work ...	.11	.31
(Co, Co, Im)	... Hello, **first time** borrow. Requesting 150$ **because** I just spent my savings on fixing my transmission on my car as well as paying for classes and recently I just paid to pay a citation fee I got. At this moment I really need this just to **help with bills & food** while classes are starting next week...	.11	.27

Table 3: Top 3 followed by Bottom 3 strategy triplets with average strategy attentions. Strategies are: **Co**ncreteness, **Re**ciprocity, **Im**pact, **Cr**edibility, and **Po**liteness. **SOS** indicates start of request; **EOS** indicates the end of a request.
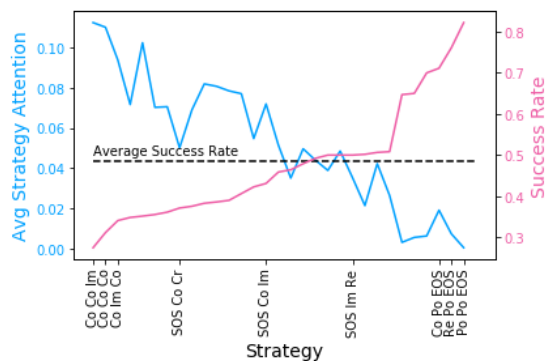

Figure 2: Strategy attention triplet vs success rate. The X-axis represents selected strategy triplets, sorted by success rate. A complete list of triplets can be found in Appendix.

actively decreased request persuasiveness, *sabotaging its success*; an under-performing strategy ordering pattern with any content often resulted in reduced persuasiveness. On the contrary, simply having an over-performing strategy with respect to the average success rate does not appear to affect a request due to reduced attention.

We also manually examined around 300 examples, with representative ones shown in Table 3. Generally, over-performing triplets had little effect on the success rates due to reduced strategy attention. However, under-performing triplets were relatively highly attended to. Below, we explain two general situations that highlight an over-performing and under-performing strategy pattern from a social science perspective:

4 Common Persuasive Patterns

4.1 "Please sir, I want some more."

A common pattern among the top 5 strategy triplets is the use of politeness. Oftentimes, the politeness

triplet appears at the end of the sentence and is usually paired with some form of reciprocity. From Figure 2, we observed that the best strategy—(Po, Po, EOS)—is a triplet with higher success rates than the average. From a social science perspective, ending a request politely engenders a sense of liking and creates connections between the audience and requester, consistent with prior work showing that politeness is a social norm associated with ending a conversation (Schegloff and Sacks, 1973). An example is shown in the first row in Table 3. However, this strategy alone does not result in a persuasive request as its associated strategy attention is relatively low. Users who end requests politely *may be likely to put effort into content,* aligning with our success rate observations. Adding to Althoff et al. (2014), we observed that users who exercise social "strategy" norms by closing conversations politely are shifting importance of a request from strategy to content. Thus, content must still be optimal for a request to be persuasive.

4.2 "It's My Money & I Need It Now."

On the contrary, if a triplet consists mostly of concreteness, it performs far below average. For instance, triplets like (Co, Co, Co) often came up in examples that were demanding as shown in Table 3. From a social science perspective, emotional appeal in arguments is key to framing aspects of a request and helps soften attention placed on facts (Walton, 1992; Macagno and Walton, 2014; Oraby et al., 2015). In the context of our dataset—a lending platform where concreteness consists mostly of demands—a lack of emotive argumentation may cause an audience to focus on demands themselves, resulting in concrete and emotionless requests.

5 Improving Request Persuasiveness

5.1 Editing Operations

Based on the effectiveness of different persuasion patterns we discovered, this section examines improving underperforming persuasion requests by editing the persuasion strategy patterns. Here we define three editing operations: (1) **Insert** (Ebrahimi et al., 2018; Wallace et al., 2019) overperforming triplets into the end of less persuasive requests containing underperforming triplets. (2) **Delete** (Ebrahimi et al., 2018) the underperforming triplets in less persuasive requests. (3) **Swap** (Ebrahimi et al., 2018) underperforming triplets by first deleting the underperforming triplets and then appending overperforming triplets to the request. We also noticed that overperforming triplets were general conversation closers; their insertion at the end of requests would not alter the intent of a message. We performed editing operations to the least persuasive requests (974 examples) containing the bottom 3 underperforming triplets—(Co, Im, Co), (Co, Co, Co), (Co, Co, Im)—using a randomly sampled triplet from the top 3 overperforming triplets (775 examples)—(Po, Po, EOS), (Re, Po, EOS), (Co, Po, EOS). Table 4 summarizes the results.

5.2 Editing Results

For **Insertion**, the underperforming requests did not improve by simply inserting a good ending triplet, partially because the underperforming request already consisted of a *sabotaging strategy*; furthermore, audiences are likely to generalize impressions from an underperforming strategy to the entire request (Ambady and Rosenthal, 1992).

Deletion of the poor strategy triplets boosted the persuasiveness of a request by mitigating the sabotaging effects of a non-persuasive strategy; however, since the content of the remaining request is mainly unedited, these request still have lower success rates than naturally occurring triplets.

Swapping the underperforming triplets with effective strategy triplets generated similar persuasiveness to deletion, suggesting again that the presence of an overperforming strategy triplet *does not* improve the persuasiveness of a request (unlike the sabotaging nature of an underperforming triplet); instead, it signals that a given request naturally contains good content since users who put effort into following social norms will likely work hard on the content. This may explain why requests that naturally contain overperforming triplets have higher

	Insert	Delete	Swap
Predicted Average Success Rate	.11	.43	.42
Δ from Original Request	+.00	+.32	+.31

Table 4: Predicted success rate and the average attention after our editing operations (μ over 30 runs).

success rates than our edited examples. Simply swapping strategies does not improve the content, and thus the persuasiveness, to a similar extent.

6 Conclusion & Future Work

In this work, we highlight important strategy orderings for request persuasiveness, and surface complex relationships between content and strategy at different request success rates. Finally, we notice improvements in persuasiveness by editing underperforming strategies. For future work, we plan to explore different techniques for explainability other than attention and compare effective strategies beyond the triplet level across different datasets. We also aim to look at the presence of different strategies across multi-modal settings; does introducing a new modal affect how effective/ineffective strategies are expressed? Furthermore, we plan to identify and compare different strategies across domains, as our work is limited to lending platforms— we expect that different domains would highlight different strategies.

References

Tim Althoff, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. 2014. How to ask for a favor: A case study on the success of altruistic requests. In *Eighth International AAAI Conference on Weblogs and Social Media*.

Nalini Ambady and Robert Rosenthal. 1992. Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological bulletin*, 111(2):256.

Larry M Bartels. 2006. Priming and persuasion in presidential campaigns. *Capturing campaign effects*, 1:78–114.

Timothy Borchers and Heather Hundley. 2018. *Rhetorical theory: An introduction*. Waveland Press.

Shelly Chaiken. 1980. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of personality and social psychology*, 39(5):752.

Marcus Tullius Cicero. 1862. *Cicero on Oratory and Orators*. HG Bohn.

Roy Peter Clark. 2016. *Writing Tools: 55 Essential Strategies for Every Writer*. Little, Brown, and Co.

Esin Durmus and Claire Cardie. 2018. Exploring the role of prior beliefs for argument persuasion. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1035–1045, New Orleans, Louisiana. Association for Computational Linguistics.

Esin Durmus and Claire Cardie. 2019. A corpus for modeling user and language effects in argumentation on online debating. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 602–607, Florence, Italy. Association for Computational Linguistics.

Javid Ebrahimi, Anyi Rao, Daniel Lowd, and Dejing Dou. 2018. HotFlip: White-box adversarial examples for text classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 31–36, Melbourne, Australia. Association for Computational Linguistics.

Lyneyve Finch. 2000. Psychological propaganda: The war of ideas on ideas during the first half of the twentieth century. *Armed Forces & Society*, 26(3):367–386.

Ivan Habernal and Iryna Gurevych. 2016a. What makes a convincing argument? empirical analysis and detecting attributes of convincingness in web argumentation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1214–1223, Austin, Texas. Association for Computational Linguistics.

Ivan Habernal and Iryna Gurevych. 2016b. Which argument is more convincing? analyzing and predicting convincingness of web arguments using bidirectional LSTM. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1589–1599, Berlin, Germany. Association for Computational Linguistics.

Ivan Habernal and Iryna Gurevych. 2017. Argumentation mining in user-generated web discourse. *Computational Linguistics*, 43(1):125–179.

Carl Iver Hovland, Irving Lester Janis, and Harold H Kelley. 1953. Communication and persuasion.

Eric Jang, Shixiang Gu, and Ben Poole. 2016. Categorical reparameterization with gumbel-softmax. *arXiv preprint arXiv:1611.01144*.

Diederik P Kingma and Max Welling. 2014. Auto-encoding variational bayes. In *Proceedings of the 2nd International Conference on Learning Representations*, Banff, Canada. International Conference on Learning Representations.

Fabrizio Macagno and Douglas Walton. 2014. *Emotive language in argumentation*. Cambridge University Press.

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

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Gaku Morio, Ryo Egawa, and Katsuhide Fujita. 2019. Revealing and predicting online persuasion strategy with elementary units. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6274–6279, Hong Kong, China. Association for Computational Linguistics.

Shereen Oraby, Lena Reed, Ryan Compton, Ellen Riloff, Marilyn Walker, and Steve Whittaker. 2015. And that's a fact: Distinguishing factual and emotional argumentation in online dialogue. In *Proceedings of the 2nd Workshop on Argumentation Mining*, pages 116–126, Denver, CO. Association for Computational Linguistics.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Samuel L Popkin and Samuel L Popkin. 1994. *The reasoning voter: Communication and persuasion in presidential campaigns*. University of Chicago Press.

Emanuel A Schegloff and Harvey Sacks. 1973. Opening up closings. *Semiotica*, 8(4):289–327.

Kumar Bhargav Srinivasan, Cristian Danescu-Niculescu-Mizil, Lillian Lee, and Chenhao Tan. 2019. Content removal as a moderation strategy: Compliance and other outcomes in the changemyview community. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW).

Chenhao Tan, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2016. Winning arguments. *Proceedings of the 25th International Conference on World Wide Web - WWW '16*.

Eric Wallace, Shi Feng, Nikhil Kandpal, Matt Gardner, and Sameer Singh. 2019. Universal adversarial triggers for attacking and analyzing NLP. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2153–2162, Hong Kong, China. Association for Computational Linguistics.

Douglas N. Walton. 1992. *The place of emotion in argument*. Pennsylvania State University Press.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, R'emi Louf, Morgan Funtowicz, and Jamie Brew. 2019. Huggingface's transformers: State-of-the-art natural language processing. *ArXiv*, abs/1910.03771.

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, Minneapolis, Minnesota. Association for Computational Linguistics.

Diyi Yang and Robert E Kraut. 2017. Persuading teammates to give: Systematic versus heuristic cues for soliciting loans. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW):1–21.

Zichao Yang, Zhiting Hu, Ruslan Salakhutdinov, and Taylor Berg-Kirkpatrick. 2017. Improved variational autoencoders for text modeling using dilated convolutions. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 3881–3890, International Convention Centre, Sydney, Australia. PMLR.

Model	Macro F-1
Random	.50
Naive Bayes	.62
BERT	.64
VAE LSTM	.61

Table 5: Request label performance on Validation Set.

A   Additional Model and Training Details

For all models, hyperparameters were manually tuned using macro-averaged F-1 scores and early stopping as our selection criterion. Models were trained using an NVIDIA RTX 2080 Ti. For our optimizers, unless otherwise specified, we use AdamW with learning rate 1e-3, betas (0.9, 0.999), eps 1e-08, and weight decay 0.01. We used Py-Torch (Paszke et al., 2019) and HuggingFace (Wolf et al., 2019) for any deep learning work.

VAE + LSTM: We minimize the following objective function for our graphical model:

$$\mathbb{E}_{\mathbf{l}\sim q(\mathbf{l}|\mathbf{x})}\left[\mathbb{E}_{\mathbf{z}\sim q(\mathbf{z}|\mathbf{x},\mathbf{l})}[\log p(\mathbf{x}|\mathbf{z},\mathbf{l})]\right.$$
$$-\mathrm{KL}[q(\mathbf{z}|\mathbf{x},\mathbf{l})\|p(\mathbf{z})]]$$
$$-\mathrm{KL}[q(\mathbf{l}|\mathbf{x})\|p(\mathbf{l})]$$

We also use LSTMs for the inference $q(z|x,l)$, generative $p(x|l,z)$, and discriminative $q(l|x)$ networks for our VAE. We set the size of z to be 64. l's size is defined by the number of unique strategies in our dataset: 6. We use the reparameterization trick in Kingma and Welling (2014) to use backprop on z; and use Gumbel's softmax (Jang et al., 2016) to model l continuously. Finally, we use CBOW Word2Vec embeddings (Mikolov et al., 2013) of size 128 to learn initial word embeddings. Our VAE was trained for 100 epochs; and our LSTM was trained for 50 epochs.

BERT Baseline: For our BERT Baseline, we finetune the small BERT Base Cased model, using the AdamW optimizer with learning rate of 2e-5 and Adams epsilon of 1e-8. Our BERT model is imported from HuggingFace's transformers repository, and was finetuned for 10 epochs.

Naive Bayes Baseline: We use the Multinomial Naive Bayes model, implemented with scikit-learn (Pedregosa et al., 2011), using default parameters.

Random Baseline: We use the dummy classifier with the random setting provided in Scikit-Learn

Rank	Shorthand	Expansion	Avg. Attention	Success Rate
1	Po Po EOS	Politeness, Politeness, EOS	0.00	0.82
2	Re Po EOS	Reciprocity, Politeness, EOS	0.01	0.76
3	Co Po EOS	Concreteness, Politeness, EOS	0.02	0.71
4	Cr Po EOS	Credibility, Politeness, EOS	0.01	0.70
5	SOS Re EOS	SOS, Reciprocity, EOS	0.01	0.65
6	SOS Po EOS	SOS, Politeness, EOS	0.00	0.65
7	SOS Im EOS	SOS, Impact, EOS	0.03	0.51
8	Re Co EOS	Reciprocity, Concreteness, EOS	0.04	0.51
9	SOS Co EOS	SOS, Concreteness, EOS	0.02	0.50
10	SOS Im Re	SOS, Impact, Reciprocity	0.04	0.50
11	Co Re EOS	Concreteness, Reciprocity, EOS	0.05	0.50
12	SOS Co Po	SOS, Concreteness, Politeness	0.04	0.50
13	SOS Co Re	SOS, Concreteness, Reciprocity	0.04	0.49
14	SOS Im Co	SOS, Impact, Concreteness	0.05	0.48
15	SOS Im Po	SOS, Impact, Politeness	0.04	0.46
16	Re Co Po	Reciprocity, Concreteness, Politeness	0.05	0.46
17	SOS Co Im	SOS, Concreteness, Impact	0.07	0.43
18	Co Co EOS	Concreteness, Concreteness, EOS	0.05	0.42
19	Re Co Co	Reciprocity, Concreteness, Concreteness	0.08	0.41
20	Po Co Co	Politeness, Concreteness, Concreteness	0.08	0.39
21	Co Co Re	Concreteness, Concreteness, Reciprocity	0.08	0.39
22	Im Co Re	Impact, Concreteness, Reciprocity	0.08	0.38
23	SOS Co Co	SOS, Concreteness, Concreteness	0.07	0.38
24	SOS Co Cr	SOS, Concreteness, Credibility	0.05	0.37
25	Co Im Re	Concreteness, Impact, Reciprocity	0.07	0.36
26	Co Co Po	Concreteness, Concreteness, Politeness	0.07	0.36
27	Im Co Co	Impact, Concreteness, Concreteness	0.10	0.35
28	Co Co Cr	Concreteness, Concreteness, Credibility	0.07	0.35
29	Co Im Co	Concreteness, Impact, Concreteness	0.09	0.34
30	Co Co Co	Concreteness, Concreteness, Concreteness	0.11	0.31
31	Co Co Im	Concreteness, Concreteness, Impact	0.11	0.27

Table 6: Extended List of Strategy Triple Attentions Ranked by Corresponding Success Rates. **SOS** indicates start of request; **EOS** indicates the end of a request.

(Pedregosa et al., 2011).

Validation performance across all classifiers can be seen in Table 5.

B   Persuasion Strategy Triplets

A full ranked list of persuasion strategy triplets, along with average strategy attention and success rates can be found in Table 6.