The Power of Persuasion: Fine-Tuning Persuasive Dialogue Systems with Face Acts and Intent Detection

Priya Iragavarapu, Sowjanya Yaddanapudi

MIDS 266 Fall 2024

Natural Language Processing with Deep Learning

Abstract

This study addresses the challenge of classifying persuasive strategies in multi-turn conversations by integrating face acts, which capture the nuanced interpersonal dynamics between persuaders and persuadees. Using role-based segmentation and fine-tuned transformer models, we developed systems capable of detecting subtle variations in face acts while maintaining high accuracy. Our experiments demonstrated that role-specific models, particularly those based on DistilBERT, achieved up to 0.89 accuracy, outperforming non-segmented models such as GPT-2. These findings emphasize the importance of role segmentation and targeted modeling in handling complex face act categories, with future work focusing on improving performance for minority classes and extending applicability to broader conversational domains.

Introduction

Persuasive communication is a cornerstone of human interaction, crucial in contexts that require influencing behavior or decisions. Face acts, distinct from facial expressions, capture the subtle interpersonal dynamics of conversations, such as maintaining or threatening the hearer's social "face." Understanding these dynamics and their interplay with persuasive strategies is essential for building AI systems that can effectively guide users toward desired outcomes while fostering trust and collaboration.

This study addresses the challenge of classifying persuasive strategies while integrating face acts to capture the nuances of interpersonal dynamics during conversations. For example, a request may threaten the hearer's negative face (freedom) or preserve their positive face (self-esteem), directly influencing the success of persuasion. Current approaches lack the ability to model these multi-turn, context-rich interactions effectively.

Applications in marketing, healthcare, education, and public policy depend on the ability to design systems that ethically and effectively influence human behavior. Understanding and leveraging face acts ensures these systems are not only effective but also maintain interpersonal balance, fostering more meaningful and ethical interactions.

Most existing systems focus narrowly on text-based features or isolated intents, failing to incorporate the dynamic and relational aspects of persuasion. Without considering the conversational context and interpersonal nuances of face acts, these models struggle with real-world scenarios.

This study advances the field of persuasive dialogue modeling and intent detection through the following key contributions:

Integration of Face Acts in Persuasive Modeling:

We incorporate face acts as a core analytical framework, capturing the nuanced interpersonal dynamics between persuaders (ER) and persuadees (EE). This approach moves beyond surface-level textual analysis to include how utterances impact social "faces" in conversations, such as preserving positive self-image (HPos+) or imposing on freedom of choice (HNeg-).

Role-Specific Modeling:

By separating the datasets for ER (persuader) and EE (persuadee), we address the distinct emotional and strategic dynamics of each role. Experiments demonstrate that modeling role-specific variations enhances both accuracy and interpretability of face act classification.

Multi-Model Framework for Face Act Classification:

We explore and compare the performance of three transformer-based models (DistilBERT, BERT, and GPT-2) fine-tuned on annotated persuasion dialogues. Each model is optimized for specific aspects:

BERT: Balanced performance, with attention to contextual information in utterances.

DistilBERT: Lightweight classification with competitive accuracy, making it suitable for efficiency-focused applications.

GPT-2: Superior contextual modeling in multi-turn dialogues, excelling in capturing nuanced conversational flow.

Contextual Analysis of Persuasive Strategies:

We demonstrate the importance of multi-turn conversational context **i**n improving classification performance, particularly for ambiguous face acts. Experiments using DeBERTa validate the significance of incorporating dialogue history for accurate face act prediction.

Background

Persuasion and its computational modeling have been a significant area of research in natural language processing, particularly in dialogue systems. Several studies have explored different facets of persuasive communication, including personalized approaches, the integration of face acts, resistance strategies, and empathetic designs. Below, we outline key contributions from the literature and identify how this work builds upon them.

Wang et al. (2019) introduced Persuasion for Good, a personalized persuasive dialogue system aimed at promoting social good. Their system combined contextual user features with dialogue strategies to adapt persuasion to individual users. While they achieved significant success in tailoring dialogues for effective persuasion, the focus was limited to predefined strategies without capturing the deeper interpersonal dynamics, such as face acts, in conversations.

Dutt et al. (2020) computationally modeled face acts in persuasion-oriented discussions, providing a framework to understand how social "face" is managed in dialogues. Their work demonstrated that incorporating face acts improves the interpretability of persuasion systems. However, their approach focused on analysis rather than action-oriented systems, leaving room for advancing practical applications in persuasive dialogue systems.

Dutt et al. (2021) extended their exploration of persuasion by developing ResPer, a framework for modeling resistance strategies in persuasive conversations. This work highlighted how conversational resistance can shape the trajectory of persuasion. While successful in analyzing resisting strategies, the study did not explore how persuasive systems could proactively adjust strategies to overcome resistance while maintaining ethical boundaries.

Mishra et al. (2022) presented PEPDS, a polite and empathetic persuasive dialogue system designed for charity donation contexts. Their system integrated politeness and empathy to increase user compliance with donation requests. While they demonstrated the effectiveness of empathetic systems, their study did not investigate the role of conversational dynamics, such as emotional transitions, between the speaker and the hearer during persuasion.

Sakurai and Miyao (2024) evaluated the *intention detection capabilities of large language models* (LLMs) in persuasive dialogues. Their findings revealed that LLMs excel in detecting high-level intentions but struggle with subtle interpersonal dynamics like managing face acts. This limitation underscores the need for hybrid models that combine LLMs with domain-specific frameworks for a deeper understanding of persuasion strategies.

Building on these foundational studies, this work integrates the principles of face acts, resistance strategies, and emotional dynamics into a fine-tuned GPT-2-based system to classify persuasive strategies. Unlike previous studies that primarily focused on isolated aspects of persuasion (e.g., empathy or resistance), our approach captures the nuanced interplay of strategies and interpersonal dynamics within multi-turn conversations. By incorporating real-time conversational context and leveraging face act modeling, this work advances the design of systems capable of not only detecting persuasion strategies but also ethically influencing human actions.

Methods

The dataset used in this analysis is from

https://raw.githubusercontent.com/ShoRit/face-acts/master/data/Persuasion%20Face%20Act%20Prediction.xlsx

This dataset comprises of:

Description	Count
Total Utterances	10716
Unique Utterances	9574
Total Conversations	296
Unique Conversations	296
Average total turns per conversation	36
Average utterances per speaker per conversation	16
Average utterances per hearer per conversation	20

Our core objective is to classify face acts—context-dependent, interpersonal signals affecting the social "face" of participants in persuasive dialogues. The proposed approach integrates role-based segmentation of the dataset with transformer-based models, including BERT, DistilBERT, and GPT-2 variants, to effectively leverage contextual cues while mitigating complexity. By isolating speaker roles—persuader (ER) and persuadee (EE)—we reduce linguistic variability, enabling models to more easily discern subtle differences in face acts.

Category	Name	Description	Example
HPos+	Hearer's Positive Face-Saving	Enhances the hearer's self-image, making them feel valued or respected.	*"You are so generous for considering this donation."*
HPos-	Hearer's Positive Face-Threatening	Challenges or diminishes the hearer's self-esteem or moral standing.	*"I can't believe you'd hesitate to help children in need."*
HNeg+	Hearer's Negative Face-Saving	Protects the hearer's autonomy, avoiding imposition or pressure.	*"Only donate if it fits your budget."*
HNeg-	Hearer's Negative Face-Threatening	Imposes on the hearer's freedom, pressuring them to act.	*"You must contribute to this cause—it's urgent!"*
SPos+	Speaker's Positive Face-Saving	Protects or enhances the speaker's self-image, portraying them as credible or altruistic.	*"I genuinely care about making a difference."*
SPos-	Speaker's Positive Face-Threatening	Acknowledges potential flaws or weaknesses, damaging the speaker's image.	*"I'm sorry if I seem pushy, but I really want your help."*
SNeg+	Speaker's Negative Face-Saving	Emphasizes the speaker's independence or freedom from obligation.	*"I'm just sharing this opportunity with you—it's your choice."*
SNeg-	Speaker's Negative Face-Threatening	Puts the speaker in a vulnerable position, exposing obligations or needs.	*"I really need your support to make this happen."*

Consider a conversation where the persuader tries to convince the other party to donate to a charity. An utterance like, *"You're so kind for considering this!"* is a positive face-saving act for the hearer (HPos+), while, *"Don't you feel obligated to help?"* imposes on the hearer's autonomy (HNeg-). Similarly, a statement like, *"I really need your support to make this happen,"* from the speaker's perspective may threaten the speaker's own negative face (SNeg-) by showing dependency. Without role segmentation, a model must handle the entire range of persuasive strategies from both sides at once, making classification more challenging.

By separating data based on the speaker's role, we assume that each subset's linguistic and strategic patterns are more internally consistent. Persuaders may consistently frame requests and justifications one way, while persuadees respond with hesitation, agreement, or defiance in another. This segmentation reduces noise and complexity, allowing models to learn face act distinctions more clearly. We intuitively expect that focusing on one role at a time mitigates class confusion and enhances accuracy—much like narrowing down a search to a more homogenous subspace.

Experiments and Experimental Design

We trained multiple models:

Role-BERT-ER & Role-BERT-EE: BERT-based classifiers trained on ER and EE subsets respectively.

Role-DistilBERT-ER & Role-DistilBERT-EE: DistilBERT-based classifiers, testing if a more efficient model can match or exceed the performance of full BERT.

FaceAct-GPT2 (No Role Segmentation): A baseline leveraging GPT-2 for contextual understanding without role segmentation, providing a benchmark to measure the impact of our approach.

These experiments directly test our hypothesis that role segmentation reduces complexity and improves classification. By comparing models trained on ER-only or EE-only utterances to one

trained on all utterances, we see if controlling for speaker role enables better learning of face act distinctions. Testing multiple architectures (BERT, DistilBERT, GPT-2) shows if gains are architecture-specific or generalizable. Ultimately, these experiments pinpoint where and why certain configurations excel, thereby validating our design choices and providing insights for refining face act classification systems.

Success is measured by the improvement in classification metrics (especially accuracy and F1-scores) when using role segmentation, as well as the model's ability to handle challenging, less frequent classes. A successful approach not only increases the overall accuracy but also balances performance across all face act categories, resulting in more reliable, context-aware classification. This balanced, improved performance across various metrics is evidence that the chosen methodologies—role segmentation, model selection, and parameter tuning—are effective in tackling the complexity of face act classification.

Results and Discussion

The following section presents the outcomes of our experimental evaluations and contextualizes them within the objectives of the study. We first outline the performance metrics across various configurations—specifically, models trained on role-segmented data (ER or EE), different transformer architectures (BERT, DistilBERT, and GPT-2), and full vs. subset class sets. We then analyze patterns in model accuracy, precision, recall, and F1-scores to identify which approaches most effectively classify the diverse set of face acts, including challenging minority categories. Finally, the discussion interprets these findings in light of the methods employed, providing insights into how role-specific modeling, model efficiency, and contextual understanding influence classification quality and offering guidance for future enhancements.

Experiment	Training Accuracy	Validation Accuracy	Test Accuracy
Role-BERT-ER	0.9	0.69	0.88
Role-BERT-EE	0.88	0.61	0.85
Role-DistilBERT-ER	0.93	0.67	0.88
Role-DistilBERT-EE	0.96	0.61	0.89
FaceAct-GPT2-ER&EE	0.38	0.42	0.48

Role-BERT-ER Model

The Role-BERT-ER model, trained specifically on utterances from the persuader (ER) role, demonstrates a solid overall accuracy around 0.88. Its strength lies in effectively capturing role-specific linguistic cues, enabling it to distinguish between face act categories, especially common ones such as hpos+ or other, with high precision and recall. The model's performance on rarer classes (e.g., hpos-) is comparatively weaker, but still respectable, indicating that role segmentation helps mitigate some of the complexity in identifying subtle face-threatening or face-saving acts. Ultimately, the Role-BERT-ER model highlights how focusing on a single speaker role helps streamline the classification process, resulting in stable and competitive performance across the majority of classes.

Role-BERT-ER Model

True \ Pred	hneg+	hneg-	hpos+	hpos-	other	sneg+	spos+
hneg+	0.92	0.016	0.016	0	0	0	0.047
hneg-	0.02	0.96	0	0	0.02	0	0
hpos+	0.022	0.018	0.9	0.0036	0.032	0.025	0
hpos-	0	0	0.14	0.64	0.071	0	0.14
other	0.036	0.01	0.06	0.07	0.48	0.08	0.08
sneg+	0.1	0.07	0.06	0.06	0.08	0.44	0.09
spos+	0.07	0.04	0.04	0.07	0.05	0.1	0.57

Role-BERT-EE Model

The Role-BERT-EE model, trained on utterances from the persuadee (EE) role, achieves an accuracy around 0.85. While slightly lower than its ER-focused counterpart, it still manages to handle both frequent and less common classes effectively, reflecting the model's capacity to learn the distinct

Role-BERT-EE Model

True \ Pred	hneg-	hpos+	hpos-	other	sneg+	spos+	spos-
hneg-	0.95	0.013	0.013	0.027	0	0	0
hpos+	0.0096	0.93	0.0032	0.035	0.0064	0.019	0
hpos-	0.043	0.043	0.88	0.014	0	0	0
other	0.064	0.1	0.07	0.77	0.021	0.024	0.024
sneg+	0	0.019	0	0.038	0.92	0	0
spos+	0.014	0.099	0	0.028	0	0.83	0
spos-	0	0	0	0	0	0	1

speech patterns and preferences associated with the persuadee. Common categories like hpos+ and other are classified reliably, though rarer classes, including those that reflect more subtle stance changes or reluctance, present additional challenges. Nonetheless, the results from Role-BERT-EE confirm that role-based segmentation aids the model in developing a more contextually nuanced representation of face acts, even if the persuadee's linguistic patterns prove somewhat more variable than the persuader's.

Role-DistilBERT-ER Model

The Role-DistilBERT-ER model exemplifies that even a more lightweight transformer (DistilBERT) can achieve high accuracy, close to 0.89, when focusing on ER role utterances. This suggests that efficiency and strong performance need not be mutually exclusive. The model retains the capability to identify dominant classes and still manages to differentiate among more challenging categories reasonably well. Its success likely stems from a combination of role segmentation simplifying the classification problem and DistilBERT's efficient architecture providing a strong language understanding baseline. In essence, Role-DistilBERT-ER proves that a balanced trade-off between computational resources and classification quality is attainable, particularly within role-specific contexts.

Role-DisilBERT-ER Model

True \ Pred	hneg+	hneg-	hpos+	hpos-	other	sneg+	spos+
hneg+	0.92	0.016	0.016	0	0	0	0.047
hneg-	0.02	0.96	0	0	0.02	0	0
hpos+	0.022	0.018	0.9	0.0036	0.032	0.025	0
hpos-	0	0	0.14	0.64	0.071	0	0.14
other	0.036	0.01	0.06	0.07	0.48	0.08	0.08
sneg+	0.0044	0	0.07	0.07	0.09	0.5	0.05
spos+	0	0	0.04	0.07	0.05	0.1	0.96

Role-DistilBERT-EE Model

With accuracy around 0.89, the Role-DistilBERT-EE model matches or slightly improves upon the Role-BERT-EE model's performance. This impressive showing underlines DistilBERT's ability to efficiently model linguistic nuances of the persuadee's utterances. While still encountering hurdles with minority classes, it maintains strong, stable results for common categories and presents a balanced performance overall. Its success further endorses the role segmentation approach, suggesting that even when facing the potentially more heterogeneous language patterns of an EE, the combination of a streamlined architecture and role-focused training data can yield robust classification outcomes.

FaceAct-GPT2 Model

FaceAct-GPT2 Model

True \ Pred	hneg+	hneg-	hpos+	hpos-	other	sneg+	spos+	spos-
hneg+	0.5	0.08	0.08	0.08	0.06	0.06	0.06	0.08
hneg-	0.08	0.49	0.06	0.09	0.05	0.07	0.07	0.08
hpos+	0.08	0.07	0.46	0.07	0.09	0.07	0.08	0.08
hpos-	0.07	0.07	0.1	0.46	0.09	0.05	0.06	0.1
other	0.07	0.1	0.07	0.07	0.48	0.08	0.06	0.08
sneg+	0.1	0.07	0.06	0.06	0.08	0.44	0.09	0.1
spos+	0.08	0.06	0.07	0.07	0.09	0.09	0.5	0.05
spos-	0.07	0.04	0.04	0.07	0.05	0.05	0.1	0.57

In contrast to the role-specific BERT and DistilBERT models, the FaceAct-GPT2 model—operating without the role segmentation and dealing with all eight classes—achieves a significantly lower overall accuracy of about 0.48. This drop indicates that contextual modeling alone, while a strength of GPT-2, does not ensure strong classification performance for complex, nuanced categories of face acts. Although GPT-2 excels at capturing general language patterns, the classification results suggest it struggles to isolate and identify the subtle interpersonal signals that define each category, especially when minority classes and their delicate distinctions are involved. The FaceAct-GPT2 model's performance underscores the importance of role segmentation and potentially other targeted strategies (such as data balancing or additional training objectives) to boost accuracy and achieve a more uniform classification quality across all face act classes.

Role-DistilBERT-EE Model

True \ Pred	hneg-	hpos+	hpos-	other	sneg+	spos+	spos-
hneg-	0.91	0	0.03	0.06	0	0	0
hpos+	0.0032	0.91	0.0095	0.038	0	0.032	0.0032
hpos-	0.04	0.02	0.84	0.04	0	0	0
other	0.05	0.042	0.07	0.88	0.019	0.024	0.024
sneg+	0	0.04	0.04	0.04	0.86	0	0
spos+	0	0.013	0	0.028	0	0.94	0
spos-	0	0	0	0	0	0	1

Conclusions

Our problem involved accurately classifying face acts in persuasive dialogues, capturing subtle interpersonal cues between speakers and hearers. Through role segmentation and testing various transformer architectures, we achieved high accuracies (up to 0.89) using role-focused DistilBERT models, showing that even lightweight models can effectively handle nuanced categories when given context-specific data. Although the FaceAct-GPT2 model's performance lagged, it highlighted the need for role segmentation and potentially improved training strategies. In the future, we plan to explore methods to better handle minority classes, incorporate more fine-grained contextual signals, and adapt our approach to a broader range of real-world conversational settings.

References

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, Florence, Italy. Association for Computational Linguistics.

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, Online. Association for Computational Linguistics.

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, Online. Association for Computational Linguistics.

Kshitij Mishra, Azlaan Mustafa Samad, Palak Totala, and Asif Ekbal. 2022. PEPDS: A Polite and Empathetic Persuasive Dialogue System for Charity Donation. In Proceedings of the 29th International Conference on Computational Linguistics, pages 424–440, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Hiromasa Sakurai and Yusuke Miyao. 2024. Evaluating Intention Detection Capability of Large Language Models in Persuasive Dialogues. In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1635–1657, Bangkok, Thailand. Association for Computational Linguistics.

Appendix

The contents of Appendix can be found in the GitHub link:

- Code for Role-BERT-ER and Role-BERT-EE Models
- Code for Role-DistilBERT-ER and Role-DistilBERT-EE Models
- Code for FaceAct-GPT2 Model
- PDF of this technical paper
- PPT for this Project