Decoding Power Dynamics: Analyzing Gen-Z Slang in Email Communication

Team : Samskrutham

Revanth (2022101049)
`naga.revanth@students.iiit.ac.in`
Keshav (2022101123)
`karthikeya.busam@students.iiit.ac.in`
Shivam (2023201019)
`shivam.chaturvedi@students.iiit.ac.in`

February 17, 2025

Problem Statement

This project addresses the challenge of identifying power relations in emails containing Gen-Z slang. Traditional models, trained on formal corpora, struggle to interpret contemporary informal language, leading to inaccuracies in power dynamics classification.

Scope

The project focuses on email communication analysis, excluding other formats such as social media or instant messaging. It includes generating synthetic emails with Gen-Z slang and comparing model performance in identifying power relations. The study is limited to English language communication for a focused analysis.

Literature Review

Linguistic Evolution

A decade ago, emails predominantly used formal business lexicon (e.g., "Please find attached"). Today, they increasingly incorporate Gen-Z vernacular and internet-influenced syntax (e.g., emojis, meme references). Traditional power relation models, trained on formal corpora [Lam, Xu, Kong, and Prabhakaran (2018)],[Raut, Chawhan, Joshi, and Kasle (2020)] , struggle to understand these emerging informal patterns.

Core Problem

This creates a mismatch where state-of-the-art (SOTA) models fail to accurately identify power relations in emails containing contemporary informal language, despite underlying hierarchical realities.

Dataset Feasibility

The dataset includes the <mark>Enron Email Corpus</mark>[Agarwal, Omuya, Harnly, and Rambow (2012)] and <mark>synthetic Gen-Z emails</mark> (Gen-Z slangs). The synthetic data will be appended to the Enron corpus to create a comprehensive dataset for analysis.

Methodology

Data Collection

- Existing datasets: Enron Email Dataset (formal)

- Gen-Z email corpus creation

Annotation Framework

- Power relation labels (e.g., manager-subordinate, peer-peer)

- Gen-Z slang identification

Model Pipeline

- Baseline: Pretrained SOTA models (BERT/RoBERTa) without Gen-Z data or some new model

- Experimental: Fine-tuned models with Gen-Z corpus

Comparative Analysis

- Compare performance of SOTA model trained without Gen-Z data and Our model.

- Compare performance of SOTA model trained with Gen-Z data and Our model.

Evaluation Strategy

The project utilizes standard classification metrics:

1. **Accuracy:** Proportion of correctly classified instances.

2. **Precision:** Proportion of true positive predictions out of all positive predictions.

3. **Recall:** Proportion of true positive predictions out of all actual positive instances.

4. **F1-Score:** Harmonic mean of precision and recall.

Timeline

- **Week 1 - 2:** Literature review on power relations.

- **Week 3 - 4:** Generating synthetic Gen-Z email data.

- **Week 5 - 6:** Implementing SOTA models and developing improvements.

- **Week 7 - 8:** Comparative analysis and report writing.

References

Agarwal, A., Omuya, A., Harnly, A., & Rambow, O. (2012, July). A comprehensive gold standard for the Enron organizational hierarchy. In H. Li, C.-Y. Lin, M. Osborne, G. G. Lee, & J. C. Park (Eds.), *Proceedings of the 50th annual meeting of the association for computational linguistics (volume 2: Short papers)* (pp. 161–165). Jeju Island, Korea: Association for Computational Linguistics. Retrieved from `https://aclanthology.org/P12-2032/`

Lam, M., Xu, C., Kong, A., & Prabhakaran, V. (2018). *Power networks: A novel neural architecture to predict power relations.* Retrieved from `https://arxiv.org/abs/1807.06557`

Raut, P., Chawhan, R., Joshi, T., & Kasle, P. (2020). Classification of power relations based on email exchange. *2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON)*, 486-489. Retrieved from `https://api.semanticscholar.org/CorpusID:226266514`