



Cracking the Empathy Code: Unveiling Empathy Patterns in Reddit Conversations

Alireza Ziabari, Gabriel A. León, Jean Luo, Kevin Kapadia, Minjeong Ko, Nuan Wen, Yi Zhang, and Zhivar Sourati



Motivation

- Empathy is important for successful social support (Zaki, 2020). When do people show empathy when providing support to others?
- This depends on the trait of the support-provider (Davis, 1983), but may also depend on how *how the support seeker asks for help*.

We ask: **what linguistic features predict the amount of empathy a support-seeking post receives?**

To answer this question, we focus on support-seeking posts on Reddit:

- In recent years, Reddit has become a platform on which people receive mental health social support (Choudhury & De, 2014)
- In prior work, readability of a post title, author of a post, and other characteristics of a post unrelated to its content predict post popularity on Reddit (Segall & Zamoshch, 2012; Glenski & Weninger, 2017).
- However, posts differ in the amount of social support they receive. The ways in which posters express distress and the topics they discuss predict the amount of condolence expressions they receive (Zhou & Jurgens, 2020). The current work extends this work by comparing three types of empathy expression in responses.

Data Structure

Data from 25 advice seeking subreddits

2857 Original Posts

Variables Included:
LIWC Scores
Number of Comments
Number of Upvotes
Topic Modeling
DDR Morality Scores

85903 Level 1 replies

Variables Included:
LIWC Scores
Number of Comments
Number of Upvotes
Topic Modeling
Language Style Matching
Empathy reaction, interpretation, and exploration

Annotation Scheme

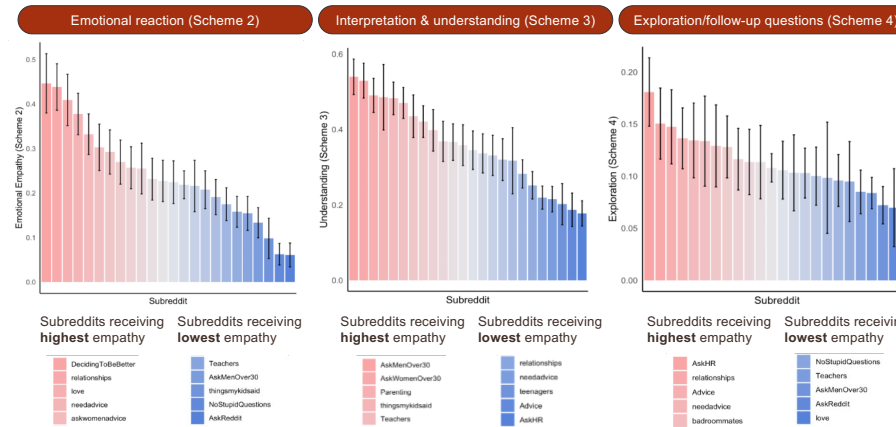
Replies were annotated for the following empathy definitions

- Emotional Reactions** - Does the response express or allude to warmth, compassion, concern or similar feelings of the responder towards the seeker?
e.g. "I'm so sorry for your loss."
- Interpretations** - Does the response communicate an understanding of the seeker's experiences and feelings? In what manner?
e.g. "I've had a similar experience before."
- Explorations** - Does the response make an attempt to explore the seeker's experiences and feelings?
e.g. "Can you tell me more about it?"

Classifying unlabeled data

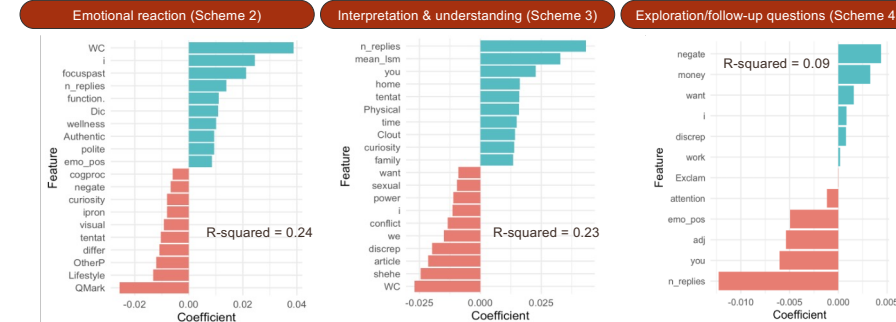
We trained a bi-encoder model based on RoBERTa (Liu et al., 2019) following the architecture proposed in Sharma et al. (2020). We obtained 3k reddit post-reply pairs with labels from the Sharma's repository and treat it as our training set. As the original labels in this dataset are No, Weak, Strong, we post-process and binarize them into No and Yes to match our annotation style. After training on this dataset, we evaluate the model on our annotated set of 494 samples. The accuracy scores are 0.71, 0.69, and 0.95 for emotional reactions, interpretations and explorations, respectively.

Do certain subreddits elicit more empathy than others?



Which linguistic features in the original post predict empathic responses in its replies?

Elastic Net Models: Empathy Score ~ LIWC Features + Word Count + Number of Replies + Mean LSM



Do empathic replies elicit more engagement?

Number of upvotes as DV		Number of comments as DV	
Independent Variables	R-Squared	Independent Variables	R-Squared
Within Post: Mean Empathy Reaction Score Between Post: Subreddit	91.73%	Within Post: Mean Empathy Reaction Score Between Post: Subreddit	96.39%
Within Post: Mean Empathy Interpretation Score Between Post: Subreddit	90.54%	Within Post: Mean Empathy Interpretation Score Between Post: Subreddit	87.55%
Within Post: Mean Empathy Exploration Score Between Post: Subreddit	89.34%	Within Post: Mean Empathy Exploration Score Between Post: Subreddit	86.06%
Within Post: Mean Language Style Matching Score Between Post: Subreddit	89.28%	Within Post: Mean Language Style Matching Score Between Post: Subreddit	86.05%

Which topics elicit more empathic responses?

After cleaning and filtering short documents, we apply Latent Dirichlet Allocation (LDA; Blei et al., 2003) for topic modeling, assuming each document is a mix of topics. We choose 10 topics after checking multiple numbers to avoid overlap. This process improves the accuracy and interpretability of the results, allowing us to extract meaningful insights from the dataset.



R-squared of Topic Modeling Predicting Empathy Scores using MLM

Topic	Empathy Reaction	Empathy Interpretation	Empathy Exploration
Topic 0	17.02%	17.26%	2.16%
Topic 1	16.47%	16.48%	2.08%
Topic 2	16.42%	18.35%	2.32%
Topic 3	16.09%	17.97%	2.52%
Topic 4	16.04%	16.89%	2.16%
Topic 5	18.69%	17.84%	2.15%
Topic 6	17.06%	16.28%	2.10%
Topic 7	16.69%	16.39%	2.10%
Topic 8	16.04%	16.61%	2.10%
Topic 9	16.26%	16.31%	2.10%

Important Findings

- Linguistic features of the original posts reliably predicted the amount of emotional reaction and understanding they receive, but not the amount of exploration responses.
- Topics such as relationships, family, and self-help/ self-actualization tended to elicit greater emotional reactions.
- Subreddits dedicated to specific social identities (e.g., women/men over 30, parents, teachers, etc.) tended to elicit responses that included greater self-disclosure of responders with similar experiences (i.e., empathy understanding).
- In general, empathy is a strong predictor of subreddit post engagement, as measured by number of upvotes and comments.