Exploring Self-Identified Counseling Expertise in Online Support Forums

Allison Lahnala, Yuntian Zhao, Charles Welch, Jonathan K. Kummerfeld,
Lawrence An, Kenneth Resnicow, Rada Mihalcea, Veronica Perez-Rosas

ACL Findings 2021

Review

Ana Sabina Uban

Contributions

Analyze social media data (Reddit) containing mental health advice, and try to determine if there is a difference between specialists and non-specialists.

Study the role of expertise in responses provided to help-seeking posts regarding mental health. Look at differences between:

(1) interactions with peers;

(2) interactions with self-identified mental health professionals (MPH)

Collect a new dataset, covering many different disorders

Overview & Goal

Motivation:

(RQ1) Do experts have distinct influences as compared to non-experts in their interactions with supportseekers in online mental health?

(RQ2) Do the experts' behaviors reflect established counseling principles and findings regarding behaviors associated with positive counseling outcomes?

Contributions

Several analyses:

- classifying between the two groups
- looking into the differences in the language used from both a semantic and stylistic perspective (based on LIWC categories complemented by other methods such as a language model for finding word importance, and a linguistic style matching score metric)
- looking further into what kind of language elicits replies received from the original support seekers.

Key findings

Interesting findings on the difference between mental health professionals (MHP) and non-professionals:

- responses from MPH vs peers can be automatically distinguished (70% accuracy)
- MHPs elicit more responses from the support seeker
- specific similar and distinctive features of MHP responses and those of high-scoring responses in general (e.g. MHPs focusing more on the second person and the future, and peers focusing more on the first person and the past).

Data Collection

They build and annotate a data of conversations between support seekers and: peers vs MHP; collected from Reddit

Many different mental health problems included (trauma, anxiety, compulsive disorders, mood disorders, addiction, eating disorders, neurodevelopmental disorders, +general health)

Self-identified MHP, but reliable high-quality annotations for MHP status: users must prove credentials. (/r/psychotherapy subreddit)

Data Collection

Post:	u/peer_user_X
I've recently been struggling with paranoid thoughts, for which I was hospitalized for my own safety. I do not feel suicidal anymore, however everyday is a long struggle of thinking everyone is an undercover agent out to get me or keep tabs on what I'm doing. I was hoping to hear some tips and stories if anyone else has dealt with similar thoughts and overcome them? Or are they something I will have to deal with for the rest of my life? Thanks in advance	
Comment:	u/MHP_user LPC
Paranoid thoughts are scared thoughts, justified or not. If you ignore the specific content of the thoughts and focus on the emotional valence (scared), is there something you can do in those moments to feel safer?	
Poster Reply:	u/peer_user_X
That's a good way of thinking about the situations as they arise. I will try to do that	

Table 2: Example of an initial post, a reply from an MHP with the flair *LPC* (Licensed Professional Counselor), and a reply from the original user.

Data Collection

Subreddits	77
Posts	12,140
Poster Replies	24,357
MHPs	283
Peers	56,701
Comments	
MHP	9,685
Peer	92,698
Total	102,383
Thread Length	
Mean	8.4
Median	4
Max	64

Table 1: Dataset statistics.

Distinguishing MHPs and Peers

Classification between responses to support seekers from MDH and peers.

Features:

a. Using as features unigrams with count > 5
b. Using as features counts of words in each LIWC category
c. Using as features counts of words in selected LIWC categories, related to perspective shifts

Model: Naive Bayes (outperformed SVM and logistic regression)

Results:

Top model: b, achieves 70% accuracy (above 50% random chance baseline)

Linguistic and Dialogue Analysis

Features used: lexicon categories in LIWC and WordNetAffect.

Methods: Find which categories are dominant for each group. (prevalence in peer group divided by prevalence in MPH

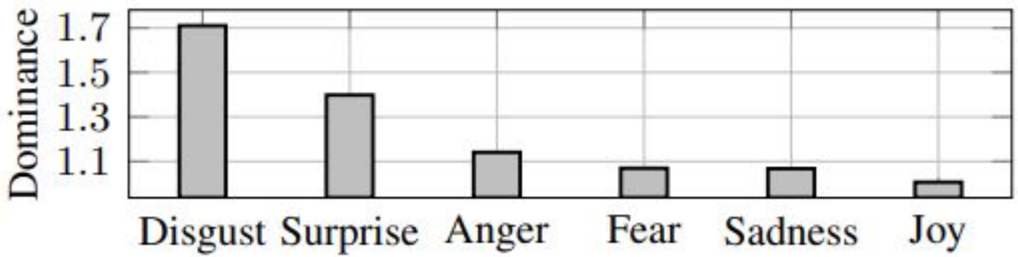


Figure 2: WordNet Affect usage (peers / MHPs)

Linguistic and Dialogue Analysis

Findings:

- more swear words and internet speak in non-professionals.
- the dominant use of first-person pronouns in peers (*I*, *we*) and focus on the past (*focuspast*). MHPs seem to use more non-first person pronouns (*you*, *they*) and focus on the future (*focusfuture*) instead.
- alignment with psychology: “clients crisis counseling conversations were more likely to report feeling better after the encounter if they exhibited perspective shifts from these categories to their counterparts (i.e., toward focusfuture, non-first person pronouns, and positive sentiment)”

Linguistic and Dialogue Analysis

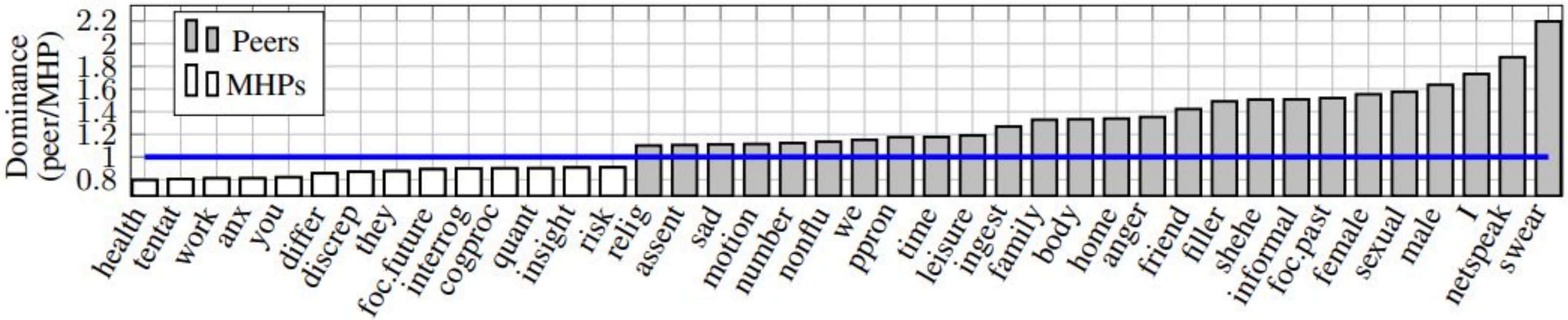


Figure 1: LIWC category dominance scores, computed as the relative use by peers divided by the relative use by MHPs, so that equal use is at $y = 1$ (blue line), higher dominance by peers at $y > 1$ (grey bars) and higher dominance by MHPs at $y < 1$ (white bars). Showing categories where frequency of use differs by at least 10%.

Engaging Support Seekers

Find dominant features in responses that prompt replies from support seekers:

- *health*, *tentat*, and *you* in the MHP comments prompting poster-replies,
- *you*, *focusfuture*, *interrog*, and *health* in the peer comments prompting poster-replies
- comments not prompting replies are similar across the two groups

Per user analysis (support giver level): dominance between replied and non-replied, quantified with Kendall-Tau coefficient:

- comments prompting engagement show similar patterns to MPH average responses. (for both groups)

Linguistic Style Matching

LSM measures the extent to which one speaker matches another. It was shown to occur more in high quality psychotherapy sessions.

Method: They measure an LSM score defined as: for words in function word categories (LIWC), difference between occurrence in post (p) and reply (r).

$$LSM_c = 1 - \frac{abs(cat\%_p - cat\%_r)}{cat\%_p + cat\%_r + .0001}$$

Linguistic Style Matching

Results:

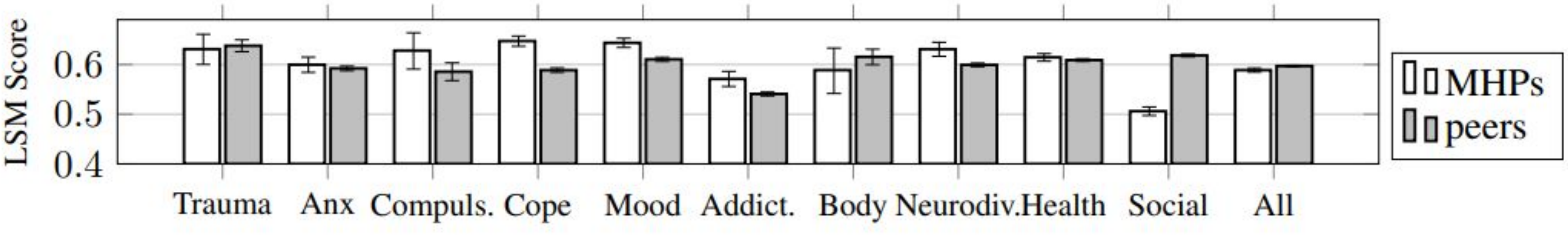


Figure 4: LSM scores with 95% confidence intervals calculated with non-parametric bootstrap resampling.

Conclusions:

- MPH may show more matching for specialized topics
- results require further investigation

Language Modelling

Assessing difference in word usage between the two groups by use of language models.

Method: Selecting words with high difference in perplexity in the two groups using the language model trained on one of the groups.

LSTM-based language model.

Words finally grouped by LIWC category.

Language Modelling

Results:

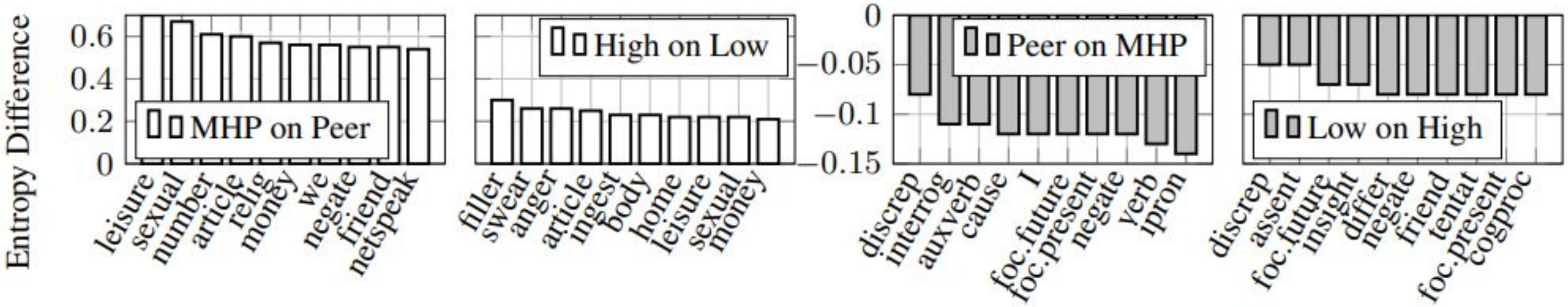


Figure 5: Entropy differences for LIWC word categories when running both language models on one group's data. High entropy scores on one dataset indicate word types that are harder for the opposite group's model to predict.

Other Strengths

Extensive review of previous work for each section, that puts study into context.

Well written and clear.

Includes extensive discussion and ethical analysis.

Detailed descriptions of experimental decisions (hyperparameters, etc) = reproducible.

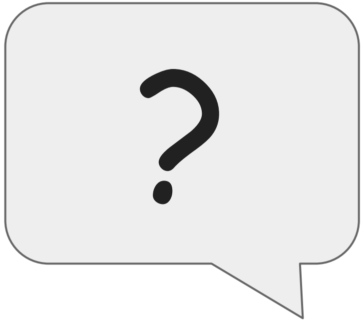
Analyses and hypotheses rely on ideas in psychology (e.g. the idea that certain language which professionals learn to use elicits more responses).

Limitations

All experiments rely on LIWC data: words in this lexicon - might miss higher-level linguistic phenomena beyond word unigram distribution.

Classification models are fairly simple.

Suggestion for **future work**: could use more sophisticated semantic representation models (language models, embeddings), or even word n-grams, to capture more information.



- What do you think?
- What did you like?
- What did you not understand?
- What could have beed done differently / what would you do as a next step?

Let’s vote!