A Computational Approach to Understanding Empathy Expressed in Text-based Mental Health Support

Ashish Sharma  Adam Miner  Dave Atkins  Tim Althoff
This presentation contains examples related to suicidal ideation and self-harm
Mental Health: Need vs. Access

Access to mental health care is poor across the globe

- Low-income, middle-income countries
  - 1 psychiatrist per 100k individuals
- United States
  - 60% of the counties do not have a single psychiatrist

Key: We may never have enough mental health professionals to meet the need

Online peer support platforms can help!
- E.g. TalkLife, 7cups, Mental health subreddits
- Millions of users seek and provide support through conversations
Text-based, asynchronous conversations on peer-support platforms

Seeker: My whole family hates me

Response post: Try talking to your friends

Peer Supporter: If that happened to me, I would feel really isolated

Communication of Empathy in Conversation

For online mental health platforms to be helpful, peer-supporters should provide effective support
Empathy

- **Core intervention component** of therapeutic counseling
- Quantitative evidence shows **strong associations with positive counseling outcomes** (Bohart et al., 2002; Elliot et al., 2011)
  - Symptom improvement
  - Alliance and rapport

"I know exactly how you feel."
Empathy: Current limitations and challenges

- Computational methods are limited to face-to-face, speech-based therapy (Gibson et al., 2016, Perez-Rosas et al., 2017).

- Previous NLP research focuses on empathy as reacting with emotions of warmth and compassion (Buechel et al., 2018) or as emotionally-grounded conversations (Rashkin et al., 2019).

- Communicating cognitive understanding of feelings and experiences of others is key in mental-health support (Selman, 1980).
Empathy expressed in Text-based Mental Health Support

- A new empathy framework and dataset
- Computational model for identifying empathy
- Model-based insights into mental health platforms

Our analysis suggests that highly empathic conversations are rare in text-based mental health support.
A new framework and dataset of empathy expressed in text-based mental health support

Adapted most prominent empathy scales from psychology/psychotherapy research to text-based mental health support in collaboration with clinical psychologists
Framework of empathy expressed in conversations

Three communication mechanisms of empathy

- **Emotional Reactions**: communicating the emotions experienced after reading a post
- **Interpretations**: communicating understanding of the inferred feelings / experiences
- **Explorations**: improving one’s understanding by exploring feelings / experiences

We differentiate between
- peers **not** expressing them at all (level 0)
- peers expressing them to some **weak** degree (level 1)
- peers expressing them **strongly** (level 2)
Tasks and Dataset

**Task 1: Empathy Identification**
How empathic is response post in the context of seeker post?

- Emotional Reactions – 1 out of 2
- Interpretations – 1 out of 2
- Explorations – 0 out of 2

**Task 2: Rationale Extraction**
What is the supporting rationale for the identified empathy levels?

Seeker: My whole family hates me. I don't see any point in living.

Supporter: I understand how you feel. Let me know if you want to talk. Everything will be fine.
Tasks and Dataset

Task 1: Empathy Identification
How empathic is response post in the context of seeker post?

- Emotional Reactions – 1 out of 2
- Interpretations – 1 out of 2
- Explorations – 0 out of 2

Task 2: Rationale Extraction
What is the supporting rationale for the identified empathy levels?
Tasks and Dataset

**Task 1: Empathy Identification**
How empathic is response post in the context of seeker post?

**Task 2: Rationale Extraction**
What is the supporting rationale for the identified empathy levels?

- Dataset of 10k (post, response) pairs annotated on our framework of empathy with supportive evidences (*rationales*)
  - 7k from TalkLife, 3k from mental health subreddits
  - Hired and trained freelancers on Upwork
  - Series of phone calls and manual/automated feedback on sample posts
  - Kappa = **0.6865**
Model for identifying empathy with supportive rationales

Multi-task, RoBERTa-based bi-encoder model
Our Computational Model

- Linear layer for input-level and token-level predictions
- Attention between the two encodings
- Two independently pretrained RoBERTa-based encoders
- Tokenize seeker post and response post

Multi-task, RoBERTa based bi-encoder model

1. Empathy identification
2. Rationale extraction

Life sucks! I lost my job
I understand how you feel

(Seeker post, response post)
Results

- Our approach identifies empathy with \(~80\%\) acc., \(~70\%\) f1
  - Random baseline would be 33\% accurate

<table>
<thead>
<tr>
<th>Model</th>
<th>Emotional Reactions acc.</th>
<th>Emotional Reactions f1</th>
<th>Interpretations acc.</th>
<th>Interpretations f1</th>
<th>Explorations acc.</th>
<th>Explorations f1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log. Reg.</td>
<td>58.02</td>
<td>51.58</td>
<td>55.53</td>
<td>41.19</td>
<td>63.23</td>
<td>51.97</td>
</tr>
<tr>
<td>RNN</td>
<td>69.09</td>
<td>54.02</td>
<td>82.25</td>
<td>47.94</td>
<td>73.40</td>
<td>28.22</td>
</tr>
<tr>
<td>HRED</td>
<td>78.91</td>
<td>48.70</td>
<td>79.26</td>
<td>29.48</td>
<td>73.40</td>
<td>28.22</td>
</tr>
<tr>
<td>BERT</td>
<td>76.98</td>
<td>70.31</td>
<td>85.06</td>
<td>62.24</td>
<td>85.87</td>
<td>71.56</td>
</tr>
<tr>
<td>GPT-2</td>
<td>76.89</td>
<td>70.76</td>
<td>80.00</td>
<td>58.43</td>
<td>83.25</td>
<td>65.65</td>
</tr>
<tr>
<td>DiatoGPT</td>
<td>76.71</td>
<td>70.42</td>
<td>85.67</td>
<td>66.60</td>
<td>83.95</td>
<td>66.34</td>
</tr>
<tr>
<td>RoBERTa</td>
<td>78.28</td>
<td>71.06</td>
<td>86.25</td>
<td>62.69</td>
<td>85.79</td>
<td>71.83</td>
</tr>
<tr>
<td><strong>Our Model</strong></td>
<td><strong>79.93</strong></td>
<td><strong>74.29</strong></td>
<td><strong>87.50</strong></td>
<td><strong>67.46</strong></td>
<td><strong>86.92</strong></td>
<td><strong>73.47</strong></td>
</tr>
</tbody>
</table>

Empathy Identification Task

- Low f1-scores, unable to distinguish between the three empathy levels
- Bi-encoder architecture works better than concatenating seeker post and response post
Results

- Both token-level (T-f1) and span-level (IOU-f1) evaluation for the rationale extraction task

<table>
<thead>
<tr>
<th>Model</th>
<th>Emotional Reactions</th>
<th>Interpretations</th>
<th>Explorations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T-f1</td>
<td>IOU</td>
<td>T-f1</td>
</tr>
<tr>
<td>Log. Reg.</td>
<td>47.44</td>
<td>63.27</td>
<td>46.92</td>
</tr>
<tr>
<td>RNN</td>
<td>62.80</td>
<td>58.22</td>
<td>67.26</td>
</tr>
<tr>
<td>HRED</td>
<td>60.56</td>
<td>55.01</td>
<td>64.26</td>
</tr>
<tr>
<td>BERT</td>
<td>61.29</td>
<td>51.20</td>
<td>61.06</td>
</tr>
<tr>
<td>GPT-2</td>
<td>47.39</td>
<td>51.27</td>
<td>64.06</td>
</tr>
<tr>
<td>DialoGPT</td>
<td>66.24</td>
<td>61.24</td>
<td>64.05</td>
</tr>
<tr>
<td>RoBERTa</td>
<td>59.12</td>
<td>63.82</td>
<td>60.08</td>
</tr>
<tr>
<td><strong>Our Model</strong></td>
<td><strong>68.49</strong></td>
<td><strong>66.82</strong></td>
<td><strong>67.81</strong></td>
</tr>
</tbody>
</table>

Rationale Extraction Task

Baselines with appropriateness to the related task of generating free-text rationales
Model-based Insights into Mental Health Platforms

Applied our model to a carefully filtered dataset of 235k interactions of significant mental health challenges on TalkLife (based on categories and triggering posts)
Validation: Positive feedback from seekers

Strong communications of emotional reactions and interpretations receive \textbf{45\% more likes} than their no communication.

Stronger explorations get \textbf{47\% more replies}.

High empathy interactions are received positively by seekers.
Validation: Forming of relationships

Seekers are 79% more likely to “follow” peer supporters after an empathic interaction than after a non-empathic one.

Relationship forming more likely after empathic interactions
How empathic are peer supporters?

- Does it improve over time?
  - This is also true for therapists!
    - Without deliberate practice and specific feedback, even trained therapists often diminish in skills over time (Goldberg et al., 2016)

Empathy expressed is low
Implications for empathy-based feedback

- We can measure empathy successfully, and the measured components are important to mental health platforms
- However, highly empathic conversations are rare
- How can you help people express more empathy?

Simple Proof-of-concept using model-based feedback

- Three participants were asked to rewrite responses using model-based feedback
- Empathy increased from 0.8 to 3
Thank You!

- Codes and dataset available at: https://github.com/behavioral-data/Empathy-Mental-Health

  @sharma_ashish_2

  [ash-shar.github.io](ash-shar.github.io)