Towards Facilitating Empathic Conversations in Online Mental Health Support: A Reinforcement Learning Approach

Ashish Sharma  Inna Lin  Adam Miner  Dave Atkins  Tim Althoff
Online Mental Health Support Platforms

**Mental health care is a global challenge**

Most countries have less than 1 psychiatrist per 100k individuals

- Widespread shortages of workforce
- Limited in-treatment options

**Online peer-support platforms can help!**

- TalkLife
- 7 CUPS
- Reddit

Millions of users **seek** and **provide support** through **conversations** online

**Seeker post**

**Response post**

Peer Supporter
Empathy

**Empathy**: The ability to understand or feel the emotions and experiences of others

High empathy interactions

- Strong associations with symptom improvement in mental health (Elliot et al., 2011)
- Received positively by users on online peer-support platforms (Sharma et al., 2020)
Highly empathic conversations are rare (Sharma et al., 2020)!

Key Question: How can we improve empathy in peer-to-peer support?
New Task: Empathic Rewriting

**Empathic Rewriting**: Computationally transform low-empathy conversational posts to higher empathy

Seeker: I can’t deal with this part of my bipolar. I need help.

Don’t worry! Try to relax. Anyone you can talk to?

Peer Supporter
New Task: Empathic Rewriting

Empathic Rewriting: Computationally transform low-empathy conversational posts to higher empathy

I can’t deal with this part of my bipolar. I need help.

Don’t worry! Try to relax. Anyone you can talk to?

Seeker

Peer Supporter
New Task: Empathic Rewriting

**Empathic Rewriting:** Computationally transform low-empathy conversational posts to higher empathy

Seeker: I can’t deal with this part of my bipolar. I need help.

Peer Supporter: Don’t worry! Try to relax. Anyone you can talk to?
New Task: **Empathic Rewriting**

**Empathic Rewriting:** Computationally transform low-empathy conversational posts to higher empathy

I can’t deal with this part of my bipolar. I need help.

Don’t worry! Try to relax. Anyone you can talk to?

Being manic is no fun. It’s scary! I’m sorry to hear this is troubling you. Try to relax. Anyone you can talk to?
New Task: Empathic Rewriting

Empathic Rewriting: Computationally transform low-empathy conversational posts to higher empathy.

Seeker

I can’t deal with this part of my bipolar. I need help.

Peer Supporter

Don’t worry! Try to relax. Anyone you can talk to?

Empathic Rewriting

Being manic is no fun. It’s scary! I’m sorry to hear this is troubling you. Try to relax. Anyone you can talk to?
Challenges: Empathy is complex

Empathy is complex, conceptually-nuanced, multi-dimensional

- Much more than sympathy or reacting with positive sentiment
- Clinically relevant perspective – Understanding of hidden feelings and experiences

Theoretically-grounded framework of empathy (Sharma et al., 2020)
Challenges: Why existing approaches fail?

- Can we transform every response to a generic, empathic response?
  - May not be **specific** to the emotions and experiences
  - Affects **response diversity** on the platform *(Althoff et al., 2016)*

- Style transfer approaches may not work!
  - Requires changes **beyond simple word-level transformations**

  **Sentiment Transfer**
  
  “The movie was bad”  ➔  “The movie was good”

  **Empathic Rewriting**
  
  “Being manic is no fun. It’s scary! I’m sorry to hear …” (3 new sentences)

- No **parallel dataset** exists and creating one is expensive
  - We will need **domain-experts**!
Empathic Rewriting: Key Contributions and Insights

We design a **Reinforcement Learning** model for empathic rewriting

- **Sentence-level** edits using **transformer LM**
- Leverage **theoretically-grounded empathy framework**
- Carefully-constructed **reward functions** for ensuring fluency, specificity, and diversity
- **Don’t need a ground truth** dataset of rewritings!

We effectively increase empathy in conversations while maintaining fluency, specificity, and diversity

35%+ more than baseline methods (BART, DialoGPT, MIME, ...)
TalkLife Dataset

- 10.9M seeker posts
- 26.9M response posts
- 642K users

Curate mental health-related conversations

- TalkLife hosts a large no. of social media interactions ("Happy mother’s day")
- Curated mental health-related conversations using a BERT-based classifier
- 1.48M seeker posts and 3.33M response posts

Computational labeling with empathy

- Label with empathy using the empathy classifier developed by Sharma et al., 2020
- Used for a supervised warm-start training of our RL model
PARTNER: Empathic Rewriting using Reinforcement Learning (RL)

PARTNER is an **RL agent** for the task of empathic rewriting.
Seeker post & Fixed-length contiguous spans of response post

PARTNER: State

Seeker post

I can’t deal with this part of my bipolar. I need help.

Response post

Don’t worry. Try to relax. Anyone you can talk to?

Seeker

Peer Supporter
PARTNER: Actions

Sentence-level edits

- Insert empathic sentences
- Replace with empathic sentences

[Action 1] Select a **position in the response span** for insertion or replacement

[Action 2] Generate candidate empathic sentences

Don’t worry. Try to relax. Anyone you can talk to?
PARTNER: Actions

Sentence-level edits
- Insert empathic sentences
- Replace with empathic sentences

[Action 1] Select a position in the response span for insertion or replacement

[Action 2] Generate candidate empathic sentences

Don’t worry. Try to relax. Anyone you can talk to?

Select the first sentence for replacement
PARTNER: Actions

Sentence-level edits
- Insert empathic sentences
- Replace with empathic sentences

[Action 1] Select a position in the response span for insertion or replacement

[Action 2] Generate candidate empathic sentences

Don’t worry. Try to relax. Anyone you can talk to?

“Being manic is no fun.”
PARTNER: Actions

Sentence-level edits
- Insert empathic sentences
- Replace with empathic sentences

[Action 1] Select a position in the response span for insertion or replacement

[Action 1] Generate candidate empathic sentences

Being manic is no fun. Try to relax. Anyone you can talk to?
Life sucks! We lost our puppy today.

I am here for you.

I know how heartbreaking this must have been. I am here for you.
PARTNER: Rewards

Ensuring highly empathic rewritings while maintaining fluency, specificity, and diversity

(1) Change in empathy
- Empathy of the rewritten response – Empathy of the original response
- Empathy scores b/w 0 to 6 based on a theoretically-grounded empathy framework (Sharma et al., 2020)

(2) Text fluency
- Measured using Perplexity of the rewritten response

\[ r_f = p_{LM} \left( \hat{R}_i \right)^{(1/N)} \]

(3) Sentence coherence
- Average sentence coherence probability between a candidate sentence and existing sentences in the response.

(4) Mutual information for specificity and diversity

\[ r_m = \lambda_{MI} \cdot \log{p} \left( \hat{R}_i|S_i \right) + \left( 1 - \lambda_{MI} \right) \cdot \log{p} \left( S_i|\hat{R}_i \right) \]
Experiments

Transformer LM initialized using pre-trained weights of DialoGPT

Warm-start supervised training using the dataset with computationally labeled empathy

Evaluation through a combination of automatic and human evaluation
Results: Automatic Evaluation

**Automatic Metrics**
- Change in empathy
- Perplexity
- Sentence coherence
- Specificity
- Diversity
- Edit rate

---

### Automatic Evaluation Metrics

<table>
<thead>
<tr>
<th>Model</th>
<th>Change in Empathy (↑)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dialogue Generation</td>
<td>0.4698</td>
</tr>
<tr>
<td>DialogGPT</td>
<td>1.2069</td>
</tr>
<tr>
<td>MIME</td>
<td></td>
</tr>
<tr>
<td>Seq-to-Seq Generation</td>
<td>0.9745</td>
</tr>
<tr>
<td>Latent Seq. BART</td>
<td>-0.0611</td>
</tr>
<tr>
<td><strong>PARTNER</strong></td>
<td><strong>1.6410</strong></td>
</tr>
</tbody>
</table>

PARTNER outperforms baselines in empathy improvement and generates fluent, specific, and diverse outputs with lower edits.

35%+ more empathy than baseline methods
We recruit six graduate students in clinical psychology with expertise in empathy and mental health support

**A/B Testing:** Compare PARTNER outputs against baseline models

Choose the output which is more (a) empathic, (b) fluent, and (c) specific
Results: Human Evaluation

PARTNER is preferred over baselines in empathy and specificity

DialoGPT
MIME
Latent Seq.
BART

Empathy
Specificity

Prefer Baselines
Prefer PARTNER
Results: Adaptability

PARTNER adapts well to different kind of responses
- Improves non-empathic responses by +2 empathy points
- Does not make an already empathic response worse
I feel like nobody cares about my existence

I feel the same

I care

It is hard to find others who can relate. I feel the same
Towards Facilitating Empathic Conversations in Online Mental Health: Summary

**Empathic conversations are** crucial for effective online mental health support, but **empathy is expressed rarely** online.

How can we **facilitate empathic conversations**?

- **New Task**: Empathic Rewriting for computationally transforming low-empathy conversations to higher empathy.
- **PARTNER**: Reinforcement Learning agent for empathic rewriting.
- **PARTNER** effectively **increases empathy** in conversations while maintaining fluency, specificity, and diversity, as demonstrated by automatic and human evaluation.

This task can be used for **giving intelligent, actionable feedback** to users with concrete, context-specific, diverse suggestions!
Thank You!

Codes and models available at: bdata.uw.edu/empathy

@sharma_ashish_2

ash-shar.github.io

Ashish Sharma  Inna Lin  Adam Miner  Dave Atkins  Tim Althoff