# It Takes Two to Empathize: One to Seek and One to Provide

**Mahshid Hosseini and Cornelia Caragea**
Computer Science Department
University of Illinois at Chicago
mhosse4@uic.edu, cornelia@uic.edu

## Abstract

Empathy describes the capacity to feel, understand, and emotionally engage with what other people are experiencing. People have recently started to turn to online health communities to seek empathetic support when they undergo difficult situations such as suffering from a life-threatening disease, while others are there to provide empathetic support to those who need it. It is, therefore, important to detect the direction of empathy expressed in natural language. Previous studies only focus on the presence of empathy at a high-level and do not distinguish the direction of empathy that is expressed in textual messages. In this paper, we take one step further in the identification of perceived empathy from text by introducing iEmpathize, a dataset of messages annotated with the direction of empathy exchanged in an online cancer network. We analyze user messages to identify the direction of empathy at a fine-grained level: seeking or providing empathy. Our dataset iEmpathize serves as a challenging benchmark for studying empathy at a fine-grained level.

## Introduction

Empathy refers to the individuals' ability to correlate with and understand others' emotional behavior and experiences (Decety and Jackson 2004). Empathy requires one to adopt the subjective viewpoint of the other (Decety and Jackson 2004) and describes the capacity to feel, understand, and emotionally engage with what other people are experiencing. Detecting and understanding empathy have applications in many domains including human-computer interaction (Buechel et al. 2018; Virvou and Katsionis 2003; De Vicente and Pain 2002), education (Virvou and Katsionis 2003), medical and healthcare (Raab 2014; Williams et al. 2015), and psychology (Yan and Tan 2014; Davis 1983; Batson 2009).

Empathy is shown to have a positive impact on creating emotional ties between people (Davis 2004). Its importance in moving patients toward a healthier state is widely discussed in the previous literature (Medeiros and Bosse 2016; Nambisan 2011; Goubert et al. 2005; LaCoursiere 2001). For example, LaCoursiere (2001) showed that when people undergo hard situations such as suffering from a severe disease, they often turn to online health communities to seek social support, of which empathy is a major component. The online communication enables users to feel empathetically supported by interacting with other people that possibly had similar experiences and can understand their feelings.

Traditional means of conveying empathy is through touch, gesture, gaze, and voice (Preece 1999). However, social networks and online communities have opened a new paradigm for expressing empathy through natural language and text-based communication. Yet with a concept as critical to augmenting patients' positive feelings as empathy (Goubert et al. 2005), to our knowledge, there has been no single computational study to detect empathetic support at a fine-grained level (i.e., seeking or providing empathy) from text. It is, therefore, important to instrument models for detecting the direction of empathy expressed in natural language.

Detecting the direction of empathy from text in the absence of visual and acoustic information is a challenging task due to: 1) the lack of annotated data; 2) the difficulty of annotating empathy from textual data due to the subjective nature of the author's intent; and 3) the empathy's sensitivity to multiple contextual circumstances. Precisely, previously studied datasets only focus on the presence of empathy at a high-level and do not make the very fine distinction between the direction of empathy that is expressed in text. However, from a psychological perspective, it is important to understand if the mood of a person seeking empathy changes after empathetic support is provided. Our models will enable such an analysis at scale. Simply detecting that a message contains empathy (without detecting the direction) is not enough to understand the impacts of empathy on people's behavior. Moreover, studying the direction of empathy is important for tasks such as empathetic response generation and can enable machines to obtain the capacity to perceive the complex affects like humans do. To our knowledge, there is no dataset available that is annotated with fine-grained empathy. In addition, in the absence of the author's intent, the perceived empathy (from the reader's perspective) may differ from the intended empathy. Furthermore, the contextual complexities such as the implicit expression of empathy and the use of metaphors need implicit reasoning, which is not accessible as surface-level lexical information. For example, the following messages *"I have* ***been in your shoes****"* and *"I am getting scared and depression is trying to* ***sneak up on me****"* lack direct surface patterns as indications of empathy, but

Copyright © 2021, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

we can infer that by the first message the author is implying "I understand you" and by the second the author is intending to say "I am about to get depression." Therefore, shallow lexical methods fail to capture deep semantic aspects.

With today's deep learning technology, the first step toward detecting empathy at a fine-grained level is to build a quality dataset. In this paper, we introduce iEMPATHIZE, a dataset compiled from an online cancer survivors' network annotated with perceived fine-grained empathy. To our knowledge, we are the first to create a dataset labeled with fine-grained empathy. We thus take one step further in the detection of empathy and identify the direction of empathetic support, seeking versus providing, from the reader's perspective. Our dataset, which is available online,[1] contains $5,007$ sentences, each annotated with a fine-grained empathy label, seeking-empathy, providing-empathy, or none.

Our contributions are three-folded: (1) We propose the task of fine-grained empathy direction detection by constructing and analyzing iEMPATHIZE, the first dataset on fine-grained empathy direction detection; (2) We establish strong baselines for a fine-grained empathy direction detection task using pre-trained language models BERT (Devlin et al. 2019). To our knowledge, this is the first work on automatically detecting the direction of empathetic support whether a message aims to provide empathy versus seek empathy. Moreover, we incorporate underlying inductive biases into BERT via domain-adaptive pre-training, which results in a better performance when integrating data from relevant domains; (3) We show that, in general, messages that provide empathy have the capacity to make a positive shift in the sentiment of participants who seek empathy.

## Related Work

There have been several studies on the importance and impacts of empathy on individuals' physiological and medical health, and its applications and benefits in various fields of psychology (Davis 1983; Batson 2009), cognitive science (Launay et al. 2015; Wakabayashi et al. 2006; Baron-Cohen and Wheelwright 2004), human-computer interaction (Virvou and Katsionis 2003; De Vicente and Pain 2002; Kort and Reilly 2002), neuroscience (Singer and Lamm 2009; Carr et al. 2003; Keysers et al. 2004), and healthcare (Williams et al. 2015; Raab 2014). Interestingly, empathy is shown to have interplay with gender, language, behavior, and culture (Chung, Chan, and Cassels 2010; Chung and Bemak 2002; Gungordu 2017). For example, in a recent study, Gungordu (2017) examined to what extent gender and cultural orientations impact people's empathetic expression and the conclusion was that women tend to show higher empathy compared to men, and individuals from different cultural backgrounds express empathy in various ways.

However, despite the importance of detecting empathy, only recently computational studies have started to be conducted on analyzing empathy from text (Sedoc et al. 2020; Yang et al. 2019a; Buechel et al. 2018; Khanpour, Caragea, and Biyani 2017; Abdul-Mageed et al. 2017) and from spoken dialogues (Pérez-Rosas et al. 2017; Fung et al. 2016; Alam, Danieli, and Riccardi 2018). Detecting empathy from a textual input in the absence of visual and acoustic information is a challenging problem for machines. Machines do not have the capacity to perceive the complex affects like humans who have the unique capability of analyzing the context and using the common-sense knowledge to communicate and express empathy. Sedoc et al. (2020) proposed a Mixed-Level Feed Forward Network and built empathy and distress lexicons consisting of $9,356$ word types along with their empathy and distress ratings. They particularly used document labels to predict word labels. Khanpour, Caragea, and Biyani (2017) proposed a model based on Convolutional Networks and Long Short Term Memory to identify empathy in health-related posts, but did not focus on the direction of empathy at a fine-grained level, which is what we do.

In an effort to identify a pathogenic type of empathy from social media language, Abdul-Mageed et al. (2017) collected a dataset of Facebook posts and assigned a pathogenic empathy score to each user based on their responses to a psychological survey. In contrast to our study, Abdul-Mageed et al. (2017) viewed empathy as a continuous variable and modeled it with a regression setup. Buechel et al. (2018) also built a corpus of people's (written) reactions to news articles rated with their (numeric) level of empathy and distress on a 7-point scale. Similar to Abdul-Mageed et al. (2017), they examined the prediction of empathy and distress as regression problems. Yang et al. (2019a) analyzed the behavioral patterns of users participating in cancer support communities and recognized eleven functional roles that members hold such as welcomer, story sharer, and support provider. They define a statistical model inspired by Gaussian mixture model (McLachlan and Basford 1988) that clusters different session representations into a set of roles. Unlike our work, Yang et al. (2019a) focused on the behavioral features of users in the online health communities and investigated how each role affects users' engagement within their communities. Wang, Zhao, and Street (2014) employed machine learning techniques to detect the different social support types in a breast cancer online community. By creating users' profiles and clustering users to different clusters, Wang, Zhao, and Street (2014) investigated how users in different groups engaged in the community. Unlike our work, Wang, Zhao, and Street (2014) used engineered features to reveal types of social support in an online health community and considered empathy as part of emotional support, not focusing on the fine-grained direction of empathy expressed in the text. De Choudhury and De (2014) studied mental health discourse on Reddit through collecting posts and comments discussing mental health issues and answered several questions on self-disclosure, social support, and anonymity. In contrast to our study, they analyzed empathy as part of emotional and social support and employed several semantic categories of words extracted from the psycholinguistic lexicon LIWC[2] to train a negative binomial regression model. Bak, Kim, and Oh (2012) employed text mining techniques to automatically analyze relationship strength and self-disclosure in Twitter chats. Balani and De Choudhury (2015) discov-

[1] https://github.com/Mahhos/Empathy

[2] http://www.liwc.net/

| | |
|---|---|
| -I cannot imagine living the rest of my life this way, I am sick to my stomach every day.<br>-Add to that the fact that I just miss him, my best friend for the past 30 years, and that's lonely. | seek |
| -I know you feel really down about this, but look at me I'm still here and have a reasonably good quality of life.<br>-I don't think you are alone in your worries I too get anxious when it is getting close for check-ups. | provide |
| -I used Aquafor skin lotion/gel (over the counter) for radiation side effects on my skin.<br>-Most insurance pays for a few counseling sessions a year, mine does. | none |

Table 1: Examples from our dataset.

ered levels of self-disclosure revealed in posts from different mental health forums on Reddit. Jaidka et al. (2020) studied online conversations from an affective perspective and introduced a conversation dataset derived from two subreddits on informational disclosure and emotional support. Yang et al. (2019b) explored how people do self-disclose in private and public channels and how these channels control the impact of self-disclosure on gaining social support.

## Dataset Construction

We introduce our new dataset for studying empathetic support direction detection called iEmpathize. Specifically, we present our dataset of $5,007$ sentences sampled from an online cancer network and annotated with three categories: seeking-empathy, providing-empathy, or none.

### Data Collection and Sampling

Our dataset is derived from an online Cancer Survivors Network[3] (CSN). CSN consists of several discussion boards attributing various topics such as different types of cancer, caregivers, emotional support, etc. In each discussion board, people can initiate a discussion thread or can comment and exchange messages in existing threads, and share their experiences, questions, and feelings about their journey with the other members. We collect data posted between 2002 and 2019 and randomly select sentences from the breast and lung cancer discussion boards for annotation. We model the empathy direction detection at *sentence level* since longer messages usually contain multiple topics, e.g., greetings, sharing everyday life events, asking for information, seeking/providing empathy, and often switch between these topics from one sentence to another. A sentence level analysis can provide a better estimation of the purpose of the author writing that sentence, whether or not expressing empathy and the empathy direction type. We filter out sentences with length less than five words. After this filtering, we obtain $5,007$ sentences in total that we use for annotation.

### Task Objective

Given a sentence, the goal is to classify it into one of the three categories: seeking-empathy, i.e., its author is perceived as seeking empathy; providing-empathy, i.e., its author is perceived as providing empathy for others, or none, i.e., its author is perceived as neither seeking nor providing empathy. Precisely, we study "perceived" fine-grained empathy, i.e., from the reader's perspective, rather than the intended empathy. Although we acknowledge that authors may have other intentions or derive other benefits of posting in forums, e.g., receiving peer support (which could be different from empathy), our motivation is that the author's intent is not always available with the text and we aim to model fine-grained empathy as perceived by a reader.

**Our Definitions** As described by Decety and Jackson (2004) empathy requires one to adopt the subjective viewpoint of the other. We define seeking empathy as asking to be truly understood, which is why people seek each other out in hard times. We define providing empathy as the psychological recognition and understanding of the others' feelings, thoughts, or attitudes. Our definitions follow the work by Decety and Jackson (2004) and several online definitions of empathy.

### Annotation and Inter-Agreement

The objective of the annotation is to label each sentence as: seeking-empathy, providing-empathy, or none. The annotation process is done in an iterative fashion and two graduate students contributed to the task. Given the difficulty of the annotation task due to its subjective nature, we purposely did not use Amazon Mechanical Turk in order to ensure the quality and consistency of annotations. Specifically, we use students who are trained internally through multiple iterations for the annotation task. Also, for the initial round of labeling, to ensure the correctness of our understanding and, therefore, the quality of the labels, we used professional advice and comments of a psychologist. In total, $1,046$ sentences are annotated as seeking-empathy, 966 as providing-empathy, and $2,995$ as none.

More precisely, we requested two graduate students to annotate the collected sentences into the three categories. We had several discussions with the annotators to explain and clarify the concept of empathy and ensure that their understanding of empathetic sentences is accurate. The annotators were also asked to read two studies of (Decety and Jackson 2004; Collins 2014) to enhance their comprehension of the task. Following prior guidelines and studies (Fort 2016; D'Mello 2015), the annotation task followed an iterative fashion. In each round, 100 sentences were assigned and disagreements were discussed with researchers. After each round of discussions, $100\%$ inter-annotator agreement (IAA) was achieved measured by Cohen's kappa coefficient. After three initial rounds of annotations, the annotators were assigned the remaining $4,707$ sentences where an IAA of $83\%$ was achieved. The last round of disagreements were discussed and labels were finalized by one of the authors of this paper. Table 1 shows examples from our dataset.

[3] https://csn.cancer.org/

**Comparison with Previous Datasets** Previous datasets for empathy detection from text (most related to our dataset) are collected from news, online cancer communities, and clinical trial studies (Buechel et al. 2018; Khanpour, Caragea, and Biyani 2017; Xiao et al. 2012). The dataset introduced by Buechel et al. (2018) has $1,860$ samples, out of which $916$ are empathetic and $944$ non-empathetic. Khanpour, Caragea, and Biyani (2017) annotated a dataset of $2,107$ samples, out of which $787$ are empathetic and $1,320$ non-empathetic. Xiao et al. (2012) introduced a dataset for empathy and obtained $854$ empathetic and $6,439$ non-empathetic utterances. In contrast, our dataset is unique in that it provides the direction of empathy (seeking or providing), rather than just detecting empathy, while having a similar or even larger size than the previous datasets.

## Characteristics of Our Dataset

In order to better understand the distinct characteristics associated with each of the classes[4] of `seek` versus `provide` empathy, we analyze the frequency of pronouns, the correlation of empathy with emotion and sentiment, and examine the most frequent noun phrases used in each class.

**Pronoun Frequency** Analyzing the frequency of pronouns in each class provides insight into the language of people seeking empathetic support and providing empathetic support for other people. As shown in Table 2, the first-person singular pronoun category has the highest frequency in the `seeking-empathy` class indicating that people who seek empathy mostly talk about themselves, their issues, and concerns. Interestingly, from Table 2, we can observe that, while the first-person singular pronoun (i.e., "I") appears frequently in the `seeking-empathy` class with $54.11\%$, it also has a frequency of $34.78\%$ in the `providing-empathy` class. On the other hand, the highest frequency in the class of `providing-empathy` belongs to the second-person (singular/plural) pronoun category, which implies that the providers' contributors intend to engage with what others are emotionally experiencing. At the same time, the second-person (singular/plural) pronoun category is the minority in the `seeking-empathy` class meaning that individuals who seek empathy rarely talk about other people in the forum. Finally, the third-person (singular) pronoun category, as can be seen in Table 2, has the second-highest frequency in the `seeking-empathy` class. This can be the indication of the cases where the user seeks empathy and shares concerns regarding a close relative's issue. For example, we observe cases like *"My dear husband now is in heaven, He was my hero, he was my everything"* or *"Lost my mom to lung cancer that Mets to the brain 8 years ago (she did NOT smoke!)"* where the users mostly talk about their close relatives.

| | | `seek` | `provide` |
|---|---|---|---|
| First person | s. | 54.11% | 34.78% |
| Second person | s./p. | 4.01% | 59.73% |
| Third person | s. | 30.49% | 19.66% |
| First person | p. | 5.25% | 7.26% |
| Third person | p. | 4.78% | 2.17% |

Table 2: Frequency percentage of pronouns per class. Singular and plural are given as s. and p., respectively.

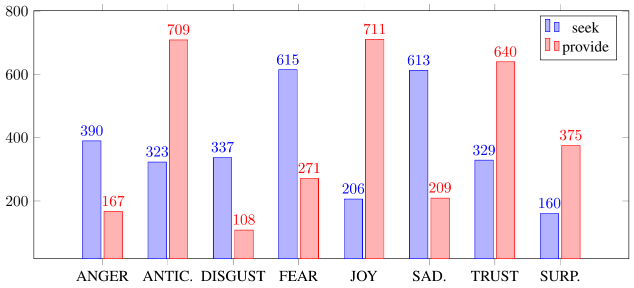

Figure 1: Fine-grained emotion co-occurrence rate per class.

**Correlation with Emotion** To understand how empathy is associated with emotion, we perform a study to assess the correlation of each empathy class with the Plutchik-8 emotion categories (Plutchik 2001), i.e. joy, sadness, anger, fear, anticipation, disgust, surprise, and trust. We particularly used the NRC emotion lexicon (Mohammad and Turney 2013), which has been widely used in different contexts such as emotion detection and sentiment analysis. The NRC emotion lexicon contains associations of English words with the Plutchik-8 emotions. Figure 1 shows the distribution of fine-grained emotions per empathy classes. One interesting observation is that fear, sadness, and anger have the most frequency in the `seeking-empathy` class indicating that people who seek empathetic support tend to utter more gloomy and depressing messages. Contrarily, when providing empathetic support, people mostly express joy, trust, and anticipation, which corroborate the fact that providers relate and understand what other people feel, and are willing to respond optimistically to other people's suffering. Having sadness, disgust, and anger the least expressed emotions in the `providing-empathy` also suggests that empathetic support is mostly provided optimistically by participants. Interestingly, results show that surprise mostly appear in the `provide` class. The observations also suggest that people rarely feel trust and joy when they seek empathetic support.

**Correlation with Sentiment** In an attempt to analyze the overall correlation of empathy and sentiment, we use the same NRC lexicon (Mohammad and Turney 2013), which also contains associations of English words with the positive and negative sentiment. The results show that the sentiment that is largely common among people seeking empathy is negative with $66\%$ frequency. Conversely, providers showed to have only $23\%$ of negative sentiment while having $77\%$ correlation with the positive sentiment. This also accords with our earlier observations, which showed that emotional feelings vary based on being a seeker or provider.

**Frequent Noun Phrases** Table 3 shows top frequent noun phrases (4-grams) in `seeking-empathy` and `providing-empathy` classes. In accordance with the

[4] We interchangeably use `seek` with `seeking-empathy` and `provide` with `providing-empathy`.

| `seek` | `provide` |
|---|---|
| i was diagnosed with | you in my prayers |
| tired of being sick | you and your family |
| i'm not sure what | will keep you in |

Table 3: Most frequent noun phrases.

| | Train | Valid | Test |
|---|---|---|---|
| | p / n | p / n | p / n |
| `seek` | 624/2,379 | 199/802 | 223/780 |
| `provide` | 584/2,419 | 182/821 | 200/801 |
| Total | 1,208/4,798 | 381/1,623 | 423/1,581 |

Table 4: Dataset splits. p/n represents positive vs. negative examples, respectively.

previous discussion, analyzing top noun phrases used by seekers and providers shows that people who seek empathetic support mostly tend to express their concerns and feelings by describing their situation. But, people who provide empathetic support for their peers tend to implicitly express their understanding and feeling towards others' pain and show their care, compassion, and support. For example, the phrase "i was diagnosed with" (Table 3) is often used by support seekers in sentences such as *"At the age of 18 I was diagnosed with lung cancer"* or *"i was diagnosed with sclc on sept 17th, 2002 and told i had approximately 30 days left"* to share their story and describe their concern. On the other hand, phrases like "you in my prayers" are mostly used by support providers. For example, in *"I hope everything goes well for you, and will remember you in my prayers"* or *"I always have all of you in my prayers but I will say an extra one for this special man in your life"* the responders tried to provide empathetic support for other users in the forum.

## Empathy Modeling

We now turn to modeling the empathy direction detection in iEmpathize. We create classifiers in two different settings. A binary setting for identifying sentences seeking empathetic support and sentences providing empathetic support. We also explore a multi-class setting for identifying sentences seeking, providing, and none (neither seeking nor providing), to compare the performance between these two different settings. To create the providing-classifier, we group the two classes of `none` and `seeking-empathy` as negative samples and keep `providing-empathy` as positive samples. Similarly, to create the seeking-classifier, we group the two classes of `none` and `providing-empathy` as negative samples and keep `seeking-empathy` as positive samples. For the second setting, we consider all three classes. We then perform a 60/20/20 split to build the train, validation, and test sets. Table 4 shows the splits. Below, we discuss our models.

**Lexicon-based Model** The NRC emotion lexicon (Mohammad and Turney 2013) and MPQA subjectivity lexicon (Stoyanov, Cardie, and Wiebe 2005) are employed to establish the baseline's feature set. We particularly used the strong and weak subjective words from the subjectivity lexicon and the words from the NRC lexicons associated with the Plutchik-8 emotions and the positive and negative sentiment. We use these as features to learn a logistic regression.

**Traditional Machine Learning** Word level TF-IDF feature vectors are extracted and employed to train various machine learning algorithms, i.e., Naïve Bayes, Support Vector Machines, and Random Forest.

**Neural Models** We conduct several experiments with a combination of different word embeddings (i.e., 300d FastText, 100d and 300d GloVe) and standard neural models: (1) **CNN:** A CNN (Kim 2014) with 100 filters of size [1, 2, 3] then are max-pooled and concatenated row-wise; (2) **ConvLSTM:** Multiple convolutional layers obtain a sequence of important features that are fed into an LSTM (Khanpour, Caragea, and Biyani 2017); (3) **LSTM:** A one-layer, unidirectional LSTM (Hochreiter and Schmidhuber 1997) with a hidden dimension of 128; (4) **Bi-LSTM:** A one-layer, bidirectional LSTM (Hochreiter and Schmidhuber 1997) with a hidden dimension of 64. The final representations from the models above are then fed into a fully connected layer and a softmax is used to obtain the final predictions. We estimate hyper-parameters on the validation set, e.g., GloVe 100d shows the best results. We report mean performance over 5 runs, each with different random initialization.

**Pre-trained Language Models** We fine-tune BERT (Devlin et al. 2019), `bert-base-uncased`, from the HuggingFace Transformers library (Wolf et al. 2020) with an added single linear layer on top as a sentence classifier that used the final hidden state corresponding to `[CLS]` token.

### Results

Table 5 presents the classification results. As we can see from the table, BERT outperforms other models in both settings. Interestingly, each setting appeared to have different levels of complexity for both the standard neural models and pre-trained language models. BERT performs remarkably well on our `provide` setting. However, our multi-class and `seek` settings remain comparably challenging for all models. The standard neural models showed to suffer from the inability to capture the complex phenomena present in our dataset with word embeddings alone. This indicates that our empathy direction prediction problem cannot be solved by solely employing pre-trained word embeddings, which do not render enough representational power to model our complex empathetic contexts. Similarly, as can be seen in Table 5, the lexical-based model comparably results in a very low F1-score in all the settings. The deficiency of sparse lexical features in the complex and implicit empathy expressions makes these methods ineffective.

## Domain-Adaptive Pre-training

To enhance our baselines, we examine domain-adaptive pre-training as a method to implicitly imbue our BERT model with the inductive biases corresponding to our domain-specific topic (Han and Eisenstein 2019; Gururangan et al. 2020). Our goal, by employing the domain-adaptive pre-training, is to optimistically make BERT more robust in the

|  | seek |  |  | provide |  |  | multi-class |  |  |
|---|---|---|---|---|---|---|---|---|---|
|  | Pr | Re | F-1 | Pr | Re | F-1 | Pr | Re | F-1 |
| Lexical-based | 63.85 | 53.51 | 51.57 | 73.58 | 58.66 | 59.66 | 60.46 | 47.24 | 48.30 |
| SVM | 76.02 | 55.28 | 53.90 | **87.92** | 75.32 | 79.28 | 75.23 | 57.22 | 58.72 |
| Naïve Bayes | **88.30** | 50.64 | 44.63 | 84.72 | 55.66 | 54.66 | **81.70** | 42.22 | 38.97 |
| Random Forest | 75.14 | 56.50 | 56.01 | 86.99 | 73.91 | 77.83 | 73.94 | 55.62 | 55.72 |
| LSTM | 38.19 | 50.00 | 43.30 | 80.48 | 76.61 | 78.25 | 69.17 | 67.15 | 68.02 |
| BiLSTM | 67.71 | 64.93 | 65.96 | 79.96 | 76.80 | 78.18 | 66.21 | 61.34 | 63.82 |
| ConvLSTM | 69.93 | 67.68 | 68.61 | 77.37 | 80.79 | 78.79 | 62.54 | 63.79 | 63.07 |
| CNN | 73.75 | 67.80 | 69.64 | 83.03 | 78.40 | 80.33 | 73.64 | 60.41 | 62.48 |
| BERT | 78.37 | **73.40** | **76.37** | 86.87 | **83.37** | **84.49** | 75.88 | **73.42** | **74.42** |

Table 5: Empathy direction identification. Best results are **bolded**.

|  | seek |  |  | provide |  |  |
|---|---|---|---|---|---|---|
|  | Pr | Re | F-1 | Pr | Re | F-1 |
| NO-PRETRAIN | 78.37 | 73.40 | 76.37 | 86.87 | 83.37 | 84.49 |
| CSN | 80.78 | 73.74 | 78.94 | 87.71 | 83.37 | 85.88 |
| filtered-CSN | 79.34 | 73.49 | 76.89 | 87.57 | 82.47 | 84.32 |
| EmoNet | 79.59 | 70.86 | 75.60 | 87.50 | 82.02 | 84.04 |
| SST | 79.77 | 67.59 | 73.98 | 83.19 | 84.72 | 83.45 |

Table 6: BERT Unsupervised intermediate task pre-training. Results are presented with green (↑) and red (↓) with respect to BERT NO-PRETRAIN on an intermediate task.

health domain and obtain helpful abstract knowledge for the end empathy direction recognition task. To this end, we first pre-train on the unsupervised tasks of masked language modeling (MLM) and next sentence prediction (NSP) (Devlin et al. 2019) on unlabeled samples from CSN, filtered-CSN (i.e., the subset of sentences from CSN that contain at least one emotion word from the NRC emotion lexicon), EmoNet (Abdul-Mageed and Ungar 2017), and SST (Socher et al. 2013), and pre-train BERT for a fixed number of epochs. Then fine-tune it on our empathy direction recognition task. Following Devlin et al. (2019), as the MLM task's prediction targets, we choose 15% of inputs at random. The corresponding targets are set to: `[MASK]` in 80% of the time, random tokens in 10% of the time, and remain unchanged in 10% of the time. Similar to MLM task, we adopted Devlin et al. (2019) settings for NSP task.

## Results

Table 6 compares the results obtained from the unsupervised intermediate task pre-training with the no intermediate pre-training setup. SST negatively impacts the performance, especially reducing the `seek` setting F1 by 3%. SST also degrades the `provide` setting F1 by 1%. The observed drop in the overall performance could be attributed to the SST's content, which mainly contains sentences in movie reviews that are incomparably distant from the health domain. The results also suggest that incorporating EmoNet yields no noticeable benefits. EmoNet covers general tweets, which are substantially different from our domain. On the other hand, pre-training on the entire CSN data slightly improves the downstream performance, particularly in the `seek` setting by 2%. Filtering CSN based on emotion words from the NRC lexicon, however, does not show a significant impact on the performance. Interestingly, it can also be seen from Table 6 that most of the settings improve the precision to some extent. Hence, we can conclude that incorporating knowledge from related tasks can be beneficial to get more relevant results (less false positives) than irrelevant ones.

## Sentiment Dynamics with Empathy

Individuals in CSN can discuss their concerns by creating a new thread on a topic and other members can respond and talk about that specific topic within the thread. As of 2019, CSN included $37,104$ unique users, $79,913$ threads, and $761,678$ posts. Table 7 shows an example of a thread where $P_{11}$ and $P_{12}$ represent originator's posts and $P_{21}$ represents a responder's post. Interacting with other people that possibly had the same (or similar) experiences and can understand their feelings makes members feel better and empathetically supported. In this section, we conduct an experiment to analyze the impact of providing empathetic support on thread originators' feelings.

The best BERT model was employed to establish the empathy label for all sentences in the CSN forum. We particularly used the `seeking` classifier to label the sentences of the initial posts (e.g., $P_{11}$ in Table 7) since intuitively these initial posts are richer in seeking empathy. We used the `providing` classifier to label the sentences posted by the responders (e.g., $P_{21}$ in Table 7). We decide a post seeks empathetic support if more than half of its sentences are labeled as `seeking` empathy. The same approach applies to the `providing` as well. In total, we tagged around 1M and 2.5M sentences by our `seeking` and `providing` classifiers, respectively. In this experiment, in total $202,717$ sentences were labeled as `seeking-empathy` and $615,638$ sentences were labeled as `providing-empathy`.

To understand whether and how the empathetic support has an impact on the feelings of the originator of a thread

| |
|---|
| ***P*$_{11}$**: *I have just finished my chemo treatments and after a brief 'high' from being finished, I found myself unsure and wondering 'what now?' Is this to be expected?* |
| ***P*$_{21}$**: *Hi, I finished chemo one year ago. I remember when my treatments were finished I experienced the same high then Oh no, what now? I kept waiting to feel like my "old" self and wondered how long till I felt better. I have to tell you HANG IN THERE! You will start to feel like your old self. One day you will wake up and realize that cancer isn't the first thought you have. Until then take care. God bless.* |
| ***P*$_{12}$**: *Hi! Thanks for responding to my message. It helps to hear from someone a year after treatment. I woke up last Sunday and for the first time in 8 months actually felt like myself!!!* |

Table 7: Illustration of change in sentiment. $P_{ij}$ denotes the $j^{th}$ post of the $i^{th}$ commentator.

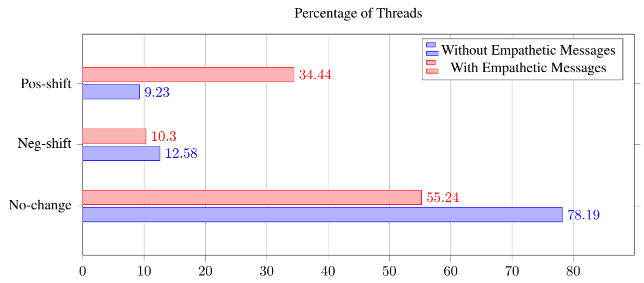

Figure 2: Thread-originator's feelings change as a result of empathetic messages in a thread. Pos and Neg represent positive and negative, respectively.

(e.g., $P_{12}$ in Table 7), we only keep the threads where the originator replied (at least) once after a comment was posted from responders. We use Stanford sentiment toolkit (Manning et al. 2014) to calculate the sentiment of originators' posts (very negative, negative, neutral, positive, very positive) and then analyze the sentiment shift by comparing the sentiment of the originator's initial post with its next posts in the corresponding thread. In total, we extracted $37,497$ discussion threads for the analysis. To recognize any changes in the originators' feelings, we grouped sentiment changes into three categories: Positive-shift, Negative-shift, and No-change. Changes towards a more positive sentiment (such as negative-to-neutral, neutral-to-positive) are represented by the Positive-shift. Negative-shift has opposite settings, and No-change exhibits cases that originator's subsequent posts echo the same sentiment as the initial post.

As shown in Figure 2, in $34.44\%$ of threads, the originators subsequently express positive-shift when they receive at least one empathetic reply from others, as opposed to only $10.3\%$ negative-shift. The results also indicate that in $55.24\%$ of the threads the originators' feelings do not change. These results signify the importance of providing empathetic support in enhancing participants' moods in online health communities. To better understand the impact of empathetic support on people's feelings, we conduct further analysis on the threads without empathetic messages in between two posts from the originator and observe that there are only $9.23\%$ positive-shift, $12.58\%$ negative-shift, and $78.19\%$ no-change. It can be therefore concluded that the participants experience positive-shift in their feelings more prominently in threads with empathetic support compared to those with no empathetic support.

## User Engagement versus Empathy

To understand the posting behaviors of most, regular, and least engaged members in terms of being a seeker or provider, we extract 20 topmost engaged members (users who posted the highest number of comments), 20 regular members (users who posted an average number of comments), and 20 bottommost engaged members (users who posted the least number of comments). We run our best BERT model, both the `seek` and `provide` settings on the users' sentences. In our analysis, we observe an interesting behavior. As shown in Figure 3, most engaged members who spend more time posting messages in CSN, mostly tend to provide empathetic support for others as compared to seeking empathetic support. On the other hand, the least engaged members do not show any significant behavior in terms of being more of a provider or seeker. We also observed that regular members similarly seek empathetic support $(18\%)$ and provide empathetic support $(17\%)$ when they are posting a comment. These observations may help to study the behavior of influential users in online health communities.

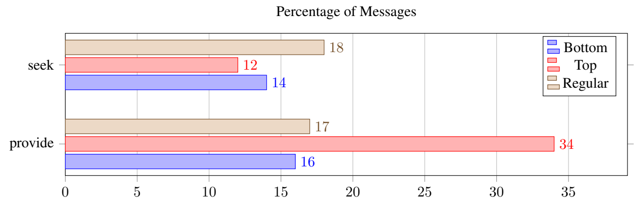

Figure 3: User Engagement versus Empathy.

## Conclusion

In this paper, we proposed the problem of fine-grained empathy direction detection from health-related community messages. To this end, we annotated a dataset, iEMPATHIZE, of messages exchanged in an online cancer network, and proposed classification tasks to distinguish fine-grained empathy direction. To the best of our knowledge, this is the first work on automatically detecting the direction of perceived empathetic support whether a message aims to provide empathy versus seek empathy. We further analyze the impact of empathetic support on peoples' feelings and show that receiving empathetic support has a positive impact on enhancing participants' moods in our studied cancer network. Identifying the direction of empathetic support considering a message's context will be part of our future work.

# References

Abdul-Mageed, M.; and Ungar, L. 2017. Emonet: Fine-grained emotion detection with gated recurrent neural networks. In *Proceedings of the 55th annual meeting of the ACL (volume 1: Long papers)*, 718–728.

Abdul-Mageed, M. M.; Buffone, A.; Peng, H.; Eichstaedt, J.; and Ungar, L. 2017. Recognizing pathogenic empathy in social media. In *11th AAAI Conference on Web and Social Media*.

Alam, F.; Danieli, M.; and Riccardi, G. 2018. Annotating and modeling empathy in spoken conversations. *Computer Speech & Language* 50: 40–61.

Bak, J.; Kim, S.; and Oh, A. 2012. Self-disclosure and relationship strength in Twitter conversations. In *Proceedings of the 50th ACL (Volume 2: Short Papers)*, 60–64.

Balani, S.; and De Choudhury, M. 2015. Detecting and characterizing mental health related self-disclosure in social media. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, 1373–1378.

Baron-Cohen, S.; and Wheelwright, S. 2004. The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *JADD* 34(2): 163–175.

Batson, C. D. 2009. These things called empathy: eight related but distinct phenomena. *MIT press* URL https://mitpress.universitypressscholarship.com/view/10.7551/mitpress/9780262012973.001.0001/upso-9780262012973-chapter-2.

Buechel, S.; Buffone, A.; Slaff, B.; Ungar, L. H.; and Sedoc, J. 2018. Modeling Empathy and Distress in Reaction to News Stories. *CoRR* abs/1808.10399. URL http://arxiv.org/abs/1808.10399.

Carr, L.; Iacoboni, M.; Dubeau, M.-C.; Mazziotta, J. C.; and Lenzi, G. L. 2003. Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proceedings of the national Academy of Sciences* 100(9): 5497–5502.

Chung, R. C.-Y.; and Bemak, F. 2002. The relationship of culture and empathy in cross-cultural counseling. *Journal of Counseling & Development* 80(2): 154–159.

Chung, W.; Chan, S.; and Cassels, T. G. 2010. The role of culture in affective empathy: Cultural and bicultural differences. *Journal of Cognition and Culture* 10(3-4): 309–326.

Collins, F. M. 2014. *The relationship between social media and empathy*. Master's thesis, Georgia Southern University. URL https://digitalcommons.georgiasouthern.edu/cgi/viewcontent.cgi?article=2189&context=etd.

Davis, M. 2004. Negotiating the border between self and other. *The social life of emotions* 19–42.

Davis, M. H. 1983. Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personality and social psychology* 44(1): 113.

De Choudhury, M.; and De, S. 2014. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *Eighth international AAAI conference on weblogs and social media*.

De Vicente, A.; and Pain, H. 2002. Informing the detection of the students' motivational state: an empirical study. In *ITS*, 933–943. Springer.

Decety, J.; and Jackson, P. L. 2004. The functional architecture of human empathy. *Behavioral and cognitive neuroscience reviews* 3(2): 71–100.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 ACL: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186. Minneapolis, Minnesota: Association for Computational Linguistics.

D'Mello, S. K. 2015. On the influence of an iterative affect annotation approach on inter-observer and self-observer reliability. *IEEE TAC* 7(2): 136–149.

Fort, K. 2016. *Collaborative Annotation for Reliable Natural Language Processing: Technical and Sociological Aspects*. John Wiley & Sons.

Fung, P.; Dey, A.; Siddique, F. B.; Lin, R.; Yang, Y.; Wan, Y.; and Chan, H. Y. R. 2016. Zara the supergirl: An empathetic personality recognition system. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, 87–91.

Goubert, L.; Craig, K. D.; Vervoort, T.; Morley, S.; Sullivan, M.; de CAC, W.; Cano, A.; and Crombez, G. 2005. Facing others in pain: the effects of empathy. *Pain* 118(3): 285–288.

Gungordu, N. 2017. *Are empathy traits associated with cultural orientation?: a cross-cultural comparison of young adults*. Ph.D. thesis, University of Alabama Libraries.

Gururangan, S.; Marasović, A.; Swayamdipta, S.; Lo, K.; Beltagy, I.; Downey, D.; and Smith, N. A. 2020. Don't Stop Pretraining: Adapt Language Models to Domains and Tasks. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 8342–8360. Online: Association for Computational Linguistics.

Han, X.; and Eisenstein, J. 2019. Unsupervised Domain Adaptation of Contextualized Embeddings for Sequence Labeling. In *Proceedings of the 2019 Conference on joint EMNLP-IJCNLP*, 4229–4239.

Hochreiter, S.; and Schmidhuber, J. 1997. LSTM can solve hard long time lag problems. In *Advances in neural information processing systems*, 473–479.

Jaidka, K.; Singh, I.; Lu, J.; Chhaya, N.; and Ungar, L. 2020. A report of the CL-Aff OffMy-Chest Shared Task: Modeling Supportiveness and Disclosure. In *Proceedings of the AAAI-20 Workshop on Affective Content Analysis*.

Keysers, C.; Wicker, B.; Gazzola, V.; Anton, J.-L.; Fogassi, L.; and Gallese, V. 2004. A touching sight: SII/PV activation during the observation and experience of touch. *Neuron* 42(2): 335–346.

Khanpour, H.; Caragea, C.; and Biyani, P. 2017. Identifying Empathetic Messages in Online Health Communities. In *Proceedings of the 8th IJCNLP*, 246–251.

Kim, Y. 2014. Convolutional Neural Networks for Sentence Classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1746–1751. Doha, Qatar: Association for Computational Linguistics. doi:10.3115/v1/D14-1181. URL https://www.aclweb.org/anthology/D14-1181.

Kort, B.; and Reilly, R. 2002. An affective module for an intelligent tutoring system. In *International Conference on Intelligent Tutoring Systems*, 955–962. Springer.

LaCoursiere, S. P. 2001. A theory of online social support. *Advances in Nursing Science* 24(1): 60–77.

Launay, J.; Pearce, E.; Wlodarski, R.; van Duijn, M.; Carney, J.; and Dunbar, R. I. 2015. Higher-order mentalising and executive functioning. *Personality and individual differences* 86: 6–14.

Manning, C. D.; Surdeanu, M.; Bauer, J.; Finkel, J. R.; Bethard, S.; and McClosky, D. 2014. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd ACL: system demonstrations*, 55–60.

McLachlan, G. J.; and Basford, K. E. 1988. *Mixture models: Inference and applications to clustering*, volume 38. M. Dekker New York.

Medeiros, L.; and Bosse, T. 2016. Empirical analysis of social support provided via social media. In *International Conference on Social Informatics*, 439–453. Springer.

Mohammad, S. M.; and Turney, P. D. 2013. Crowdsourcing a Word-Emotion Association Lexicon. *Computational Intelligence* 29(3): 436–465.

Nambisan, P. 2011. Information seeking and social support in online health communities: impact on patients' perceived empathy. *Journal of the American Medical Informatics Association* 18(3): 298–304.

Pérez-Rosas, V.; Mihalcea, R.; Resnicow, K.; Singh, S.; and An, L. 2017. Understanding and predicting empathic behavior in counseling therapy. In *Proceedings of the 55th Annual Meeting of the ACL (Volume 1: Long Papers)*, 1426–1435.

Plutchik, R. 2001. The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist* 89(4): 344–350. ISSN 00030996.

Preece, J. 1999. Empathic communities: Balancing emotional and factual communication. *Interacting with computers* 12(1): 63–77.

Raab, K. 2014. Mindfulness, self-compassion, and empathy among health care professionals: a review of the literature. *Journal of health care chaplaincy* 20(3): 95–108.

Sedoc, J.; Buechel, S.; Nachmany, Y.; Buffone, A.; and Ungar, L. 2020. Learning Word Ratings for Empathy and Distress from Document-Level User Responses 1664–1673. URL https://www.aclweb.org/anthology/2020.lrec-1.206.

Singer, T.; and Lamm, C. 2009. The social neuroscience of empathy. *Annals of the New York Academy of Sciences* 1156(1): 81–96.

Socher, R.; Perelygin, A.; Wu, J.; Chuang, J.; Manning, C. D.; Ng, A. Y.; and Potts, C. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on EMNLP*, 1631–1642.

Stoyanov, V.; Cardie, C.; and Wiebe, J. 2005. Multi-perspective question answering using the OpQA corpus. In *Proceedings of the conference on human language technology and EMNLP*, 923–930. Association for Computational Linguistics.

Virvou, M.; and Katsionis, G. 2003. Relating error diagnosis and performance characteristics for affect perception and empathy in an educational software application. In *Proceedings of the 10th International Conference on Human Computer Interaction (HCII) 2003*, 22–27.

Wakabayashi, A.; Baron-Cohen, S.; Wheelwright, S.; Goldenfeld, N.; Delaney, J.; Fine, D.; Smith, R.; and Weil, L. 2006. Development of short forms of the Empathy Quotient (EQ-Short) and the Systemizing Quotient (SQ-Short). *Personality and individual differences* 41(5): 929–940.

Wang, X.; Zhao, K.; and Street, N. 2014. Social support and user engagement in online health communities. In *International Conference on Smart Health*, 97–110. Springer.

Williams, B.; Brown, T.; McKenna, L.; Palermo, C.; Morgan, P.; Nestel, D.; Brightwell, R.; Gilbert-Hunt, S.; Stagnitti, K.; Olaussen, A.; et al. 2015. Student empathy levels across 12 medical and health professions: an interventional study. *Journal of Compassionate Health Care* 2(1): 4.

Wolf, T.; Debut, L.; Sanh, V.; Chaumond, J.; Delangue, C.; Moi, A.; Cistac, P.; Rault, T.; Louf, R.; Funtowicz, M.; Davison, J.; Shleifer, S.; von Platen, P.; Ma, C.; Jernite, Y.; Plu, J.; Xu, C.; Le Scao, T.; Gugger, S.; Drame, M.; Lhoest, Q.; and Rush, A. 2020. Transformers: State-of-the-Art Natural Language Processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Online: Association for Computational Linguistics.

Xiao, B.; Can, D.; Georgiou, P. G.; Atkins, D.; and Narayanan, S. S. 2012. Analyzing the language of therapist empathy in motivational interview based psychotherapy. In *Proceedings of The 2012 APSIPA ASC*, 1–4. IEEE.

Yan, L.; and Tan, Y. 2014. Feeling blue? Go online: an empirical study of social support among patients. *Information Systems Research* 25(4): 690–709.

Yang, D.; Kraut, R. E.; Smith, T.; Mayfield, E.; and Jurafsky, D. 2019a. Seekers, providers, welcomers, and storytellers: Modeling social roles in online health communities. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–14.

Yang, D.; Yao, Z.; Seering, J.; and Kraut, R. 2019b. The channel matters: Self-disclosure, reciprocity and social support in online cancer support groups. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–15.