**IST 736 Text Mining** **Final Report**

**Team Members: Xi Liu, Lu Guo**

**Understand the Knowledge, Attitudes, and Experiences of Palliative Care: A Natural Language Processing Study of Reddit Posts**

(All the data can be downloaded from: <https://drive.google.com/drive/folders/16JlzFbIudDpbaavuv9xaV8AITvlz-wgn?usp=drive_link>)

**Introduction**

Palliative care aims to improve seriously ill patients' quality of life. The 4th edition of Clinical Practice Guidelines for Quality Palliative Care was released by The National Consensus Project for Quality Palliative Care in 2018. The updated guideline expands the offer of palliative care to all people with serious illnesses, regardless of diagnosis, prognosis, or age (Staff, n.d.).

The study has shown that the majority of cancer survivors reported knowing no or little about palliative care (Stal et al., 2022). The results suggested a need to increase palliative care knowledge among cancer survivors to address disparities in the acceptance and utilization of it (Stal et al., 2022).

Therefore, it is necessary to understand palliative care from patients and their caregivers' perspectives since they are essential parts of the treatment. This study focuses on the knowledge, stance, practical and emotional challenges patients and caregivers encounter, and the social support they seek during palliative care. Based on the insights, prevention strategies can be developed to improve the experiences of palliative care not only for the patient's physical comfort but also for patients' and caregivers' emotional health.

Individuals with various illnesses connect with others via social media, share their experiences, provide and receive support (Vance et al., 2009). Peer-to-peer support is a kind of social support from people who have experienced similar situations (Naslund et al., 2016). The foundation of this approach is a sense of connection between peers who have similar experiences, and both parties can consequently offer more authentic empathy and validation than professional knowledge (Mead & Macneil, 2004).

There is evidence that people with diverse concerns are willing to look for peer-to-peer online support (Abbas et al., 2021). Additionally, the anonymity of online forums is found to be positive when it comes to mental illness because expressing oneself without stress can be an effective healing process (De Choudhury & De, 2014).

Therefore, this study analyzes user-generated content in online forums via natural language learning methods to investigate cancer patients and their caregivers' knowledge, stance, and experiences of palliative care. Three research questions are proposed:

RQ1: What are the concerns related to palliative care people posted on online group forums?

RQ2: What are users' attitudes/stance toward palliative care?

RQ3: Are palliative care patients and caregivers looking for and providing emotional support on online group forums?

**Method**

**Social Media Platform Choice**

This study examines submissions and comments from one of the most popular social news and entertainment media, Reddit. In 2021, roughly 430 million users worldwide registered on Reddit, making it the 5th most visited site in the U.S. Users can post content, including texts, photos, and videos. Other users can then comment on these posts. Furthermore, users can also post comments underneath a post. Therefore, there is a "thread," a series of comments, and "subthreads", threads with threads underneath a post. Users can vote positively ("upvote") or negatively ("downvote") on every post, both the initial post and subsequent posts commenting on the initial post. The Reddit system calculates the cumulation of a post's upvotes and downvotes, then the post gets a post's number, which determines its position on a web page, higher up or lower down in the list of posts displayed. This allows most people to read the most popular posts.

Reddit is organized by areas of interest or sub-communities called "subreddits" (denoted by "r/"). One distinctive aspect of Reddit is its anonymity. The Reddit "throwaway" accounts allow users to express sensitive information and uninhibited feelings. The issues related to mental health discussed on Reddit include self-disclosure of challenges in daily activities and personal relationships (De Choudhury & De, 2014). Feedback on these postings ranges from emotional and instrumental commentary to informational advice (De Choudhury & De, 2014).

**Data Collecting**

This study used Python to collect submissions, comments, and associated metadata of six cancer subreddits (testicularcancer, subredditcancer, cancer, breastcancer, braincancer, and thyroidcancer) from the top 20k subreddit dumps until 2022-12-30 (https://atlantis.ischool.syr.edu/share/reddit/cancer/). Each subreddit dump includes submissions, comments, and related metadata.

After being filtered by the word “palliative,” a total of 596 submissions and 2815 comments were left (Table 1).

*Table 1: Data distribution in six subreddits*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | testicularcancer | subredditcancer | cancer | breastcancer | braincancer | thyroidcancer | **total** |
| **submission** | 6 | 1 | 527 | 26 | 35 | 1 | **596** |
| **comment** | 21 | 2 | 2,514 | 122 | 151 | 5 | **2,815** |

Corresponding to each original submission, the researchers collected information on a) the title, b) the textual content, c) the timestamp when the post was made, d) the post's score, e) the subreddit name, f) the subreddit id, g) author name, h) author id, i) submission link, and j) the number of comments it obtained.

For each comment, information on the a) textual content, b) the timestamp when the comment was made, c) author name, d) author id, e) parent id, f) subreddit name, g) subreddit id, h) comment link, and i) the score of the comment posts are collected.

**Analysis**

First, we explore the data to see the distribution. The result shows that the discussion about palliative care significantly increased since 2018 (Image 1). The increase coincides with the release of the 4th edition of the palliative care guideline that extended the scope to all patients with serious illnesses. The result indicates that the data are appropriate for our study purpose.

A graph with blue and orange lines

Description automatically generated

*Image 1: Data distribution between 2010 to 2022.*

We use Topic Modeling to analyze the submissions and identify the concerns related to palliative care. Stance Analysis is used to assess users’ attitudes toward palliative care. For identifying emotional support in comments, we use the Classification Model.

**Results**

**RQ1: What are the concerns related to palliative care people posted on online group forums?**

To answer RQ1, we use Topic Modeling. We used sentence\_embedding+kmeans to cluster submissions and comments individually. We set the K=10 for submissions and K=7 for comments.

The Topic Modeling results of submissions indicate seven major topics: 1) emotional distress after diagnosis, 2) challenges, 3) coping with a loved one with a terminal illness, 4) seeking alternative treatments during palliative care, 5) seeking advice of receiving palliative care or not, 6) seeking emotional support, and 7) seeking people with similar experiences.

The Topic Modeling results of comments indicate five major topics: 1) palliative care medications and pain management, 2) navigating palliative and hospice care options, 3) support and decision-making, 4) palliative care experiences, and 5) emotional support.

We use keyBERT to validate the Topic Modeling results. For submissions, we look for the top 25 keywords (Image 2). After excluding the common words in cancer context (chemotherapy, chemo, cancer, oncologist, etc.), the keywords are: depressed, depression, cancerfamilysupport, antidepressant, cancercaregivers, depress, antidepressants, coping, miserable, restlessness, remission. These words indicate that patients and caregivers are in stressful conditions and seek coping strategies. The results echo the Topic Modeling results. A bar graph with blue and white text

Description automatically generated A graph with blue lines and words

Description automatically generated

*Image 2: Top 25 keywords of submissions Image 3: Top 35 keywords of comments*

For the comments, after excluding common words in the cancer context (cancer, tumor, chemotherapy, ect.), the keywords are: cancerscared, cancercaregivers, cancerversary, cancerhelp, oncologynurseadvisor, cancertopics, cancergrace, cancerouskelly (Image 3). We check the meaning of these words. Some of them are website name, such as CancerHelp, Oncology Nurese Advisor, CancerGRACEThese. Some are influencers of cancer forums, such as cancerouskelly. Some are encouraging words, for example, cancerversary, a word that means milestones deserve celebrating during the fighting cancer journey. These words suggest that comments provide information and emotional support and validate the results of Topic Modeling.

Therefore, on Reddit cancer-related subreddits, concerns related to palliative care mainly include mental health challenges, medical decision suggestion seeking, and emotional support seeking.

**RQ2: What are users’ attitudes/stance toward palliative care?**

To answer RQ2, first, two authors separately annotated 20 submissions and 30 comments that were randomly selected and reached an agreement on different annotations after discussion. One irrelevant comment is removed. We use four labels: positive, negative, neutral, and cannot tell. For submissions, 20% (n=4) are positive, 65% (n=13) are neutral, 10%(n=2) are cannot tell, and none (n=0) is negative. For comments, 62% (n=18) are positive, 3% (n=1) are negative, 24% (n=7) are neutral, and 10% (n=3) are cannot tell (Table 2).

|  |  |  |
| --- | --- | --- |
|  | submissions | comments |
| positive | 4 (20%) | 18 (62%) |
| negative | 0 | 1 (3%) |
| neutral | 13 (65%) | 7 (24%) |
| cannot tell | 2 (10%) | 3 (10%) |
| Total | 20 | 29 |

*Table 2: Annotation results of submissions*

We use ChatGPT to classify the stance. Firstly, we tried the prompt "What's the stance towards palliative care in the below submission/comment from Reddit? Answer positive, negative, neutral, or cannot tell as accurately as possible. The submission/comment is: ###{}###" The macro-F1 score is 0.39. The confusion matrix is:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | prediction | | | | |
| true |  | positive | negative | neutral | cannot tell |
| positive | 15 | 1 | 7 | 0 |
| negative | 0 | 0 | 1 | 0 |
| neutral | 9 | 2 | 8 | 1 |
| cannot tell | 1 | 0 | 2 | 2 |

Previous research has shown that using the internal label of ChatGPT can improve the macro-F1 score (Kim et al., 2023). To get a higher macro-F1 score, we try to find labels from the internal of ChatGPT. We asked ChatGPT, "What labels are usually used in a stance classification task?" ChatGPT responded to four labels "Support, Oppose, Neutral, Unrelated." Then We revised our prompt to "What's the stance towards palliative care in the below submission/comment from Reddit? Answer Support, Oppose, Neutral, or Unrelated as accurately as possible. The submission/comment is: ###{}###". The macro F-1 score decreased to 0.33.

The results show that stance classification did not work well on our data. Manual annotation experiences indicate that the submissions and comments rarely show a stance on palliative care, but the emotions and attitudes during the palliative care process.

**RQ3: Are palliative care patients and caregivers looking for and providing emotional support on online group forums?**

To answer RQ3, one author annotated 121 submissions manually first for training and evaluation purposes. The manual annotation results show that 31% (n=37) of submissions sought emotional support, and 69% (n=84) were not.

Then, we build SVM and BERT models to classify whether a submission is looking for or supporting emotional support. For SVM, we removed stopwords and used a boolean vectorizer. The macro-F1 score of SVM is 0.61, and BERT is 0.38. The SVM model performs better than the BERT model. The BERT model predicts all the content as “having no emotional support.”

**Conclusion**

First, Reddit is a good place for people to express their feelings and seek social support. The coincidence of the increase of the submissions in six subreddits with the updated palliative care guideline demonstrates that people intended to share feelings and questions with others with similar experiences.

Second, regarding palliative care, more caregivers, instead of patients, shared online. The potential explanation is that there were few cancer survivors in these six cancer subreddits to share their experiences and feelings. For future investigation, we need more data from subreddits related to other diseases that need palliative care, not only focusing on cancers.

Third, the manual annotation and the clustering results from the six subreddits suggest that caregivers encountered severe stress during palliative care. The primary sources of stress were the fear of facing a loved one dying, and the additional tasks caused by palliative care, such as preparing funnels. Caregivers were actively seeking informational, instrumental, and emotional support from those who had similar experiences. One specific usage of these subreddits was venting. Users vented their negative experiences and feelings in a group of people who had similar experiences, so they knew that they were understood and welcomed.

Last, Reddit users provided informational and emotional support via comments. The informational support included educational and medical resources, such as websites. The emotional support included encouraging words and sharing experiences.

Therefore, based on the results of this study, the caregivers of cancer patients who had palliative care need more attention when discussing the improvement of palliative care. Since caregivers are a critical part of palliative care, their well-being, especially mental health, will impact the overall experiences of palliative care. Mental health prevention strategies should be developed to support caregivers so they can better support patients, spend more time with patients with higher quality, and have more positive memories if they lose a loved one.

**Ethics Statements**

There are two major ethical concerns. The first one is the usage of user-generated content. We used the content without informing and obtaining agreement from the users who created them. Additionally, the users might be identified if we quote their content. The second ethical concern is the bias in the data. We only focus on six cancer-related subreddits; the results of this study may not apply to all diseases that need palliative care. Also, our results have potential biases due to the accuracy level of the modeling results.

**Declare:** We used ChatGPT API to classify the stance toward palliative care. The prompts and performance have been stated in the previous part.

**References**

Abbas, J., Wang, D., Su, Z., & Ziapour, A. (2021). The Role of Social Media in the Advent of COVID-19 Pandemic: Crisis Management, Mental Health Challenges and Implications. Risk Manag Healthc Policy, 14:1917–1932. [https://doi](https://doi/): 10.2147/RMHP.S284313

Choudhury, M. D., & De, S. (2014). Mental Health Discourse on reddit: Self-Disclosure, Social Support, and Anonymity. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), 71–80. <https://doi.org/10.1609/icwsm.v8i1.14526>

Kim, Y., Guo, L., Yu, B., & Li, Y. (2023, July). Can ChatGPT Understand Causal Language in Science Claims?. In Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, *Sentiment, & Social Media Analysis* (pp. 379-389).

Low, D. M., Rumker, L., Talkar, T., Torous, J., Cecchi, G., & Ghosh, S. S. (2020). Natural Language Processing Reveals Vulnerable Mental Health Support Groups and Heightened Health Anxiety on Reddit During COVID-19: Observational Study. *Journal of Medical Internet Research*, *22*(10), e22635. <https://doi.org/10.2196/22635>

Naslund, J. A., Grande, S. W., Aschbrenner, K. A., & Elwyn, G. (2014). Naturally occurring peer support through social media: the experiences of individuals with severe mental illness using YouTube. *PLOS one*, *9*(10), e110171. <https://doi.org/10.1371/journal.pone.0110171>

Mead, S., & MacNeil, C. (2006). Peer support: What makes it unique. *International Journal of Psychosocial Rehabilitation*, *10*(2), 29-37.

Staff, A. I. W. (n.d.). *Palliative care guidelines updated with increased focus on collaboration, communication*. Retrieved December 12, 2023, from https://acpinternist.org/weekly/archives/2018/11/06/4.htm

Stal, J., Nelson, M. B., Mobley, E. M., Ochoa, C. Y., Milam, J. E., Freyer, D. R., & Miller, K. A. (2022). Palliative care among adult cancer survivors: Knowledge, attitudes, and correlates. *Palliative & Supportive Care*, *20*(3), 342–347. <https://doi.org/10.1017/S1478951521000961>

Vance, K., Howe, W., & Dellavalle, R. P. (2009). Social internet sites as a source of public health information. *Dermatologic clinics*, *27*(2), 133-136. <https://doi.org/10.1016/j.det.2008.11.010>