Natural Language Processing as a tool to identify patient sentiment in GI related internet forums

Abhishek Rao[1], Shahzaib Maqbool[2], Rasheed Nassr[2] & Andrew Wilson[2]

[1] George Eliot Hospital, Gastroenterology, Birmingham, United Kingdom.
[2] College of Computing, Birmingham City University, Birmingham, United Kingdom

Introduction:
Natural language processing (NLP) refers to the ability of machines to understand and explain human language. It has potential applications for clinicians in identifying how their patients express and interact with their conditions.

Internet forums allow for exchange of detailed information and is used by patients to discuss their illnesses with those similarly affected.

This study collected data from user comments on abdominal pain in a GI forum to demonstrate the technology's potential for automated identification of patients' feelings on relevant GI topics.

Methods:
The online patient forum selected was MedHelp (https://www.medhelp.org/posts/Digestive-Disorders-Gastroenterology), chosen due to its interface being amenable to scraping tools. For this study, the investigators chose to review how "abdominal pain" was enquired within these forums.

Data was collected via an AI-powered visual web scraping tool (Scrapestorm).

This was analysed using the Natural Language Toolkit, Pandas software library for data manipulation, Scikit-learn for number classifications, regression and clustering, and Spacy for string comparisons. Collection/Counter to tally the frequency of phrases in discussions and their likelihood in the entire dataset.

Sentiment (i.e emotional tone of text) was calculated using exploratory data analysis and was graded on a three-point ordinal scale (good, negative, neutral).

Results:
6342 discussions were identified within MedHelp evaluating for the following words:
- o abdominal pain
- o lower pain
- o upper pain
- o left sided pain
- o right sided pain.

This yielded 2318 records likely related to abdominal pain.

A word cloud of the most frequently used terms identified within these is listed in Figure 1.

Figure 1. Word cloud of Trigrams

Top occurring phrases (n(4) gram) include 'I thought I going', 'feel like I need', 'I feel like I', which indicate patients' uncertainty about their condition.

Sentiment analysis of the data identified that 64.6% expressed a positive interest in the discussion, 28.6% expressed a negative reaction and 6.8% had a neutral reaction.

Conclusions:

To our knowledge, this is the first study to use NLP and sentiment analysis for GI related online forums. The implementation of this technology is still in its infancy and further refinements are essential for more comprehensive topic analysis and understanding of user sentiments.

NLP allows automated extraction and analysis of extensive data, helping clinicians better understand patient sentiments, of particular value for chronic illnesses like IBD and often misunderstood conditions such as IBS.