

Student name: Sheikh Bilal Ahmad Supervisor contact name: Dr. Conor Hayes

Student number: 24238647 Supervisor contact email: conor.hayes@universityofgalway.ie

Gauging Public Opinion on Health Topics using Knowledge Graph Summarization

Introduction

Medical Science has always been a field of utmost importance that helped people lead healthier lives and cure ailments. However, the general public may have some conceptions, even misconceptions, regarding topics that have health implications. My project aims to use NLP techniques and knowledge graphs to give medical experts an idea of public opinion on health topics. For this, I will be using the comment sections of New York Times articles on health topics as the database on which knowledge graph summaries will be generated.

Background Research or Context

To understand public opinion on health topics, the comments people make on health articles on New York Times will be used. Using text summarisation (extractive or abstractive) can be used to summarise these comments. A knowledge graph can integrate text summary elements and frequency and we can use it to generate text summaries (by ranking and selecting nodes and edges as was done for news text summarisation^[1]) or opinion frequency tables if we wish.

Proposed Project

The data collection will be done via the API provided by the New York Times. These comments are unstructured data in natural language and can be noisy. Some challenges in the processing of these comments is making sense of abbreviations and coreference resolution especially since some comments are lengthy. These articles talk about a health topic and the comments usually consist of people agreeing/disagreeing, choosing between alternative and how much they trust doctors. So, firstly, I plan on extracting the entities in the title of the article. Then apply entity recognition and relationship extraction (with coreference resolution) to the comment to extract what people say/believe about the entities in the title (key entities) and also regarding doctors. Then, use these entities and relations to construct a knowledge graph for each article. There are two kinds of information that is expected to be extracted for each key entity: positive/negative sentiment and the qualitative opinion of the commenter.

For analysing which entity is being talked about (like in the first few lines in the above comment) and for providing more context to the knowledge graph using existing knowledge, I plan to use and fine-tune a large language model like BERT. I need to do a literature review on the text analysis techniques to accomplish these NLP tasks and also on LLMs, to choose which model will suit my use-case the best.

The extracted information will be presented in the form of a dashboard. The dashboard on its homepage will present a list of topics. For each topic, a list of analysed New York Times articles with links will be present. Clicking on a topic will take the user to the page for that topic where the following panels will be present: 1. key entities and the sentiment associated with them and the number of instances for each sentiment, 2. key entities and the qualitative information extracted (as a text summary, generated with the help of an LLM) and 3. A brief visualisation of the knowledge graph generated. If an article for a topic is clicked, then these 3 elements for that article will be presented. Some quantitative information like number of comments per topic and article etc. can also be integrated in this dashboard. A search functionality will be built into the dashboard for easy navigation and querying.

I will exclusively focus on health related articles in the New York Times.

Timeline

11/04: Data collection complete

19/04: Literature review complete

03/05: Entity recognition and relationship extraction code complete

17/05: Aspect-based sentiment analysis complete code

24/05: Knowledge graph construction code complete

14/06: Using knowledge graph and LLM to generate text summary output – code complete

28/06: Dashboard complete

31/07: Project complete

References:

[1]

M. V. P. T. Lakshika, H. A. Caldera, and W. V. Welgama, "Abstractive Web News Summarization Using Knowledge Graphs," vol. 2, pp. 300–301, Nov. 2020, doi: <https://doi.org/10.1109/icter51097.2020.9325453>.