Obesity - Using NLP to determine causes, consequences, and solutions

Ron Isaac
Pace University, Seidenberg School

Git hub: https://github.com/ctronisaac/nlp-obestity-project
Notebook: https://docs.google.com/presentation/d/1BwIUqppRFd0ASNdPJnVV3T4cmmDeHgT3KuCs9idd7uQ/edit?usp=sharing

Abstract

Obesity is a worldwide phenomenon with health and cost issues that affect us all; either by being obese and having poor health or having to pay for treatments of diseases made worse by obesity, reducing the amount of money available for other medical treatments.
In food science, it is hard to know what is true. The science keeps changing, first low fat was the recommendation, then low carbs, and now intermittent fasting. How can we keep up? This project allows a person to review the data from scientific articles and find something that motivates or educates them to reduce their weight or the weight of a loved one. By turning 41 pages of five scientific papers into 40 sentences of causes, consequences, and solutions, a person can keep up to date with the scientific community.
For me, the statement that stood out and has a high likelihood of changing my behavior is, "the public at large must fully understand that obesity is a true crisis". I am risk adverse, and yet I don't fully appreciate that my extra weight is a grave risk to my health. Using NLP (Natural Language Processing) can automate knowledge acquisition from documents and will allow us to learn about the world. The output also specifies which document contained the sentence, so if desired, the user can read the entire article.

Research Question

What does the scientific literature say are the causes of obesity, the consequences, and solutions? Can NLP be used to extract the needed information from the documents?

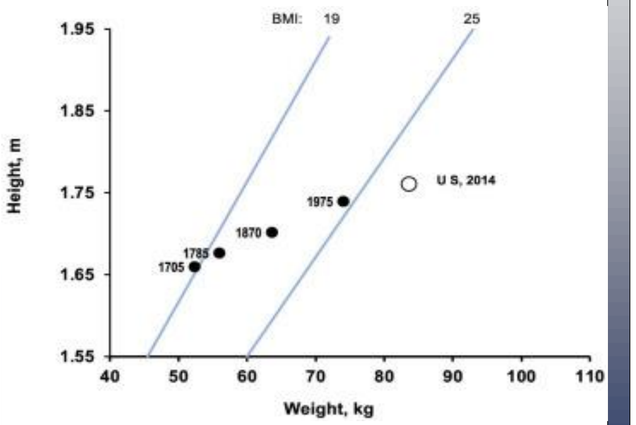
Related Work

NLP has been successfully used in categorizing medical advice [1], and also to process abstract ArXiv papers and classify them into meaningful categories [2]

Dataset

The data for this project are research documents that are then input to the NLP program I made. These documents provide research results for obesity: causes, consequences, and solutions. These research papers [3-7] were found using PubMed, which is part of the NIH (National Institutes of Health).

Increasing BMI



Methodology

The software is written in Python program utilizing libraries for reading PDFs (tika) and for NLP (spaCy). The documents of interest are placed into a folder. Each sentence in each document is examined three times. First for cause, then for consequence, and then for a solution. Each sentence that is displayed and then manually copied to a spreadsheet.

Results

The results are stored in a spreadsheet showing source document, the type (cause, solution, or consequence), and the text (the sentence). All results can be found at:
https://docs.google.com/spreadsheets/d/1dIrwLTsMdhFJNtv-PVyZJshEbyJNIhKh/edit?usp=sharing&ouid=101605668230972498483&rtpof=true&sd=true

One interesting solution was
"… the public at large must fully understand the magnitude and far-reaching effects of excess body weight and treat it as a true crisis…"

	A	B	
	Document	Type	Text
	A systematic literature revie	Cause	energy i
	A systematic literature revie	Consequece	obesity
	A systematic literature revie	Cause	combina
	A systematic literature revie	Cause	excess d
	A systematic literature revie	Cause	adverse
	A systematic literature revie	Solutions	none of
	Causes of obesity - a review	Cause	Monoge
	Causes of obesity - a review	Cause	Althoug
	Causes of obesity - a review	Solutions	ameliora

Conclusions, Limitations & Future Work

NLP is a powerful technology allowing us to use trusted sources of information to quickly obtain data to help our understanding of issues. In this case, the desire was to understand obesity causes, consequences, and solutions.

My use of NLP was primitive, and the future work will be to update the NLP models and commands used to find meaning in the documents. For example, instead of simple text matching, we can use word vectors so that similar words will also be caught.(e.g. word vectors will show that *result* is similar to *solution*)

References

1.[Yingya Li. 2022] An NLP Analysis of Health Advice Giving in the Medical Research Literature
2.[Rosanna Turrisi 2023] Beyond original Research Articles Categorization via NLP
3.[David R. Meldrum 2017] Obesity pandemic: causes, consequences, and solutions—but do we have the will?
4.[Junke Gao 2021] The Relationship of Leptin, Exercise and Dietary Habits of Overweight and Obese Children
5.[Mahmood Safaei 2021] A systematic literature review on obesity: Understanding the causes & consequences of obesity
6.[Beenish Masood 2023] Causes of obesity
7.[Benjamin Caballero 2019] Humans against Obesity: