NLP Lab 5 Report

Loic Sine
Max Tuttle

March 10, 2023

https://github.com/https://github.com/loics21/COS470—NLP

1 Introduction

In this lab, we aim to compare the answers generated from ChatGPT and those given by real users. To do this, we rely on StackExchange websites , specifically, the Computer Engineering StackExchange website.

2 Questions

For this section of the lab, we must find and choose 10 questions with varying difficulties from our chosen StackExchange website to use as our control test. To select these questions, we consider three major factors, Tags, Difficulty, and Score. Below are the 10 selected questions in order of increasing difficutly.

Question NO. 1 : "Why not put servers in a refrigerator?"
https://engineering.stackexchange.com/questions/42598/why-not-put-servers-in-a-refrigerator

Question NO. 2: "How to visualize isometric parallel projection?"
https://engineering.stackexchange.com/questions/48937/how-to-visualize-isometric-parallel-projection

Question NO. 3: "Why do we need more powerful computers nowadays?"
https://engineering.stackexchange.com/questions/49741/why-do-we-need-more-powerful-computers-nowadays

Question NO. 4: "What is the difference b/w '>>endl'; and ;?"
https://engineering.stackexchange.com/questions/49559/what-is-the-difference-b-w-endl

Question NO. 5: “What sensors/methods can be used from the roadside to determine if a truck is loaded or empty - without weighing it?”
https://engineering.stackexchange.com/questions/33511/what-sensors-methods-can-be-used-from-the-roadside-to-determine-if-a-truck-is-lo

Question NO. 6: “How to find the smallest change in voltage”
https://engineering.stackexchange.com/questions/26970/how-to-find-the-smallest-change-in-voltage

Question NO. 7: ”How can I transmit signals from pre-amplifiers to a laptop nearby using bluetooth?”
https://engineering.stackexchange.com/questions/28980/how-can-i-transmit-signals-from-pre-amplifiers-to-a-laptop-nearby-using-bluetoot

Question NO. 8: “What is the significance of a processors bit capacity?”
https://engineering.stackexchange.com/questions/7355/what-is-the-significance-of-a-processors-bit-capacity

Question NO. 9: “How did scientists develop the stencils that made even more precise processors than they had previously?”
https://engineering.stackexchange.com/questions/22444/how-did-scientists-develop-the-stencils-that-made-even-more-precise-processors-t

Question NO. 10: “What is the mechanism of transition in applications of finite state automata?”
https://engineering.stackexchange.com/questions/26101/what-is-the-mechanism-of-transition-in-applications-of-finite-state-automata

3 Analysis

Finally, we can analyze our results. For this lab, we will take analyze the results in two different ways. Firstly, we will complete a manual analysis and then an automatic analysis.

To conduct a manual analysis, we check and decide if the answers provided by real users and chatGPT are relevant to the question asked and then check to see if they are correct.

To conduct an automatic analysis, there are two types of automatic analyses we need to apply. Firstly, we calculate the overlap between the n-gram (uni-gram & bi-gram) and the vector similarity of answers. We will be relying on the OpenAI embedding models to calculate vector similarities. Finally, we compare the similarity between the embeddings using cosine similarity.

Question Num	Max Answer	Loic Answer	N-gram Result	Vector Similarity Result
1	1	1	0.0	0.68429
2	1	1	0.0	0.76428
3	1	1	0.0	0.64606
4	1	1	0.0	0.78559
5	1	1	0.0	0.71107
6	0	0	0.0	0.74112
7	1	1	0.0	0.70044
8	1	1	0.0	0.72121
9	1	1	0.0	0.70513
10	1	1	0.0	0.70372

3.1 N-Gram

Evaluating the results from n-gram probabilities, we find that the n-gram probabilities are 0.0. This result seems unlikely though.

3.2 Vector Similarities

After looking at the results from the vector similarities, there don't seem to be any glaring outliers when it comes to relatedness. The most similar answer from chatGPT and stack exchange is question 4 which discusses the difference between endl; and ; in c++. It makes sense that these answers would be the most similar because the question is not difficult and has a definitive answer. The other similarities between answers are not bad by any means considering that most of them hover around 0.7 or 70%.