Caring is Sharing – Exploiting the Value in Data for Health and Innovation M. Hägglund et al. (Eds.) © 2023 European Federation for Medical Informatics (EFMI) and IOS Press. This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0). doi:10.3233/SHTI230282

Large Language Model as Unsupervised Health Information Retriever

Keyuan JIANG^{a,1}, Mohammed M. MUJTABA^a and Gordon R. BERNARD^b ^a Purdue University Northwest, Hammond, Indiana, USA ^b Vanderbilt University, Nashville, Tennessee, USA ORCiD ID: Keyuan Jiang https://orcid.org/0000-0002-1565-3202

Abstract. Retrieving health information is a task of search for health-related information from a variety of sources. Gathering self-reported health information may help enrich the knowledge body of the disease and its symptoms. We investigated retrieving symptom mentions in COVID-19-related Twitter posts with a pretrained large language model (GPT-3) without providing any examples (zero-shot learning). We introduced a new performance measure of total match (TM) to include exact, partial and semantic matches. Our results show that the zero-shot approach is a powerful method without the need to annotate any data, and it can assist in generating instances for few-shot learning which may achieve better performance.

Keywords. Large language model, unsupervised learning, zero-shot learning, health information retrieval, COVID-19 symptoms, Twitter

1. Introduction

Recent advancements in pre-trained large language models (LLMs) have revolutionized the ways of natural language processing. They achieved state of the art (SOTA) performances in many NLP tasks including conversation, text generation, reasoning, translation, and question answering. To retrieve the symptom mentions from a corpus of COVID-19-related tweets, we treat the retrieval as a question answering task. This treatment facilitated selecting the prompt to the language model and measuring the performance of the unsupervised (zero-shot) learning.

2. Method

The 3rd generation pre-trained transformer (GPT-3) language model [1], developed by Open AI, is an autoregressive language model developed based upon the previous generation (GPT-2) [2]. We used the GPT-3 model (text-davinci-002) without modifications of its architecture and hyperparameters for zero-shot and few-shot learning (5-shot and 10-shot)².

¹ Corresponding Author: Keyuan Jiang, Purdue University Northwest, 2200 169th Street, Hammond, Indiana, U.S.A., E-mail: kjiang@pnw.edu.

² The work of the few-shot learning methods is under review at another conference.

Choosing an appropriate prompt to the language model is an important step. We adopted the prompt of "List the symptoms in this tweet as a list" after having experimented several different prompts. We proposed the total match (TM) for the practice use cases, rather than the widely used, strict exact match (EM) [3]. TM includes exact, partial and semantical matches. A corpus of 655 COVID-19-related tweets, annotated by two authors, was used to test the zero-shot learning.

3. Results and Discussions

Table 1 are the performance results of the zero-shot method along with the results of both 5-shot and 10-shot methods using GPT-3 on a corpus of 655 annotated tweets.

Table 1. Performance results of zero-shot method in comparison with 5-shot and 10-shot learning methods on 655 tweets.

	Exact Match	Total Match	Weighted TM	Precision	Recall	F1
0-shot	0.109	0.212	0.790	0.779	0.856	0.816
5-shot	0.191	0.377	0.853	0.844	0.890	0.866
10-shot	0.264	0.429	0.906	0.912	0.846	0.877

Both EM and TM are lower in our method than that of few-shot learning, suggesting that supervised learning even with a small number of examples can perform better. Precision, recall and F1 are somehow lower but comparable, showing the power of GPT-3 to recognize symptom mentions in tweets with the general domain training data. This can be useful in a larger scale health information retrieval task using few-shot learning as the zero-shot method can serve as an initial step of few-shot learning by identifying a small number of instances needed to condition the model.

It was observed that the GPT-3 language model sometimes predicts symptom mentions not found in the tweets. This may come from the facts in the data it learned. Another limitation is its inability to handle negation correctly - e.g., it predicts the phrase of "no cough" as a symptom.

4. Conclusion

The zero-shot approach of the LLM is very powerful, and can help us identify examples for few-shot learning methods when no annotated data are available, and can facilitate a larger scale task of health information retrieval.

References

- Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, Agarwal S. Language models are few-shot learners. Advances in neural information processing systems. 2020;33:1877-901.
- [2] Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Language models are unsupervised multitask learners. OpenAI blog. 2019 Feb 24;1(8):9.
- [3] Rajpurkar P, Zhang J, Lopyrev K, Liang P. Squad: 100,000+ questions for machine comprehension of text. arXiv preprint arXiv:1606.05250. 2016 Jun 16.