CovidDialog: Medical Dialogue Datasets about COVID-19

Zeqian Ju, Subrato Chakravorty, Xuehai He, Shu Chen, Xingyi Yang, Pengtao Xie

UC San Diego

PENGTAOXIE2008@GMAIL.COM

Abstract

The pandemic of COVID-19 has caused 1,345,048 infections and 74,565 deaths as of April 6th in 2020. The pandemic also causes burnout of medical professionals and panic of the public. To alleviate the burnout and panic, medical dialogue systems that are able to provide COVID-related consultations to citizens are in great need. To facilitate the research and development of such systems, we build two medical dialogue datasets – CovidDialog – containing patient-doctor conversations about COVID-19 and other related pneumonia. The datasets have an English dataset containing 603 consultations and 1232 utterances and a Chinese dataset containing 399 consultations and 8440 utterances. The datasets are publicly available at https://github.com/UCSD-AI4H/COVID-Dialogue

1. Introduction

Coronavirus disease 2019 (COVID-19) is currently spreading across the world, causing more than 800,000 infections and 40,000 deaths as of April 1st in 2020. Under the pandemic, medical resources in all countries are extremely scarce, and medical staff is facing tremendous work intensity and pressure. What's worse, fear spreads like a virus. Fearful reactions to COVID-19 can put lives at risk. For example, healthy people with mild cough symptoms go to hospital to ask a doctor for advice and get infected in the hospital, which is not only regrettable, but also increase the burden of medical work.

These circumstances motivate us to build a dialogue system that can answer COVID-19 related questions from patients. Currently, there are several online platforms where patients who are concerned that they may be infected by COVID-19 or other pneumonia can consult doctors and doctors can provide advice here. However, compared with the huge number of patients consulting doctors, the number of doctors on the online platform is far from enough. Besides, doctors cannot answer the questions 24 hours without rest. A dialogue system which can answer COVID-19 related questions from patients automatically and real-time will reduce the workload of doctors greatly. Meanwhile, people's panic will be relieved since their questions can be answered any time and their worries can be alleviated in time. Reducing medical pressure, relieving people's anxiety, and effectively assisting medical work, a dialogue system is on demand and has a promising future in application.

To facilitate the research and development of COVID19-targeted dialogue systems, we build two medical dialogue datasets that contain conversations between doctors and patients, about COVID-19 and other pneumonia: (1) an English dataset containing 603 con-

Description of medical condition

I have a little fever with no history of foreign travel or contact. What is the chance of Covid-19?

Dialogue

Patient: Hello doctor, I am suffering from coughing, throat infection from last week. At that time fever did not persist and also did not felt any chest pain. Two days later, I consulted with a doctor. He prescribed Cavidur 625, Montek LC, Ambrolite syrup and Betaline gargle solution. Since then throat infection improved and frequent cough also coming out. Coughing also improved remarkably though not completely. From yesterday onwards fever is occuring (maximum 100-degree Celcius). I have not come in touch with any foreign returned person nor went outside. In our state, there is no incidence of Covid-19. Please suggest what to do?

Doctor: Hello, I can understand your concern. In my opinion, you should get done a chest x-ray and CBC (complete blood count). If both these are normal then no need to worry much. I hope this helps.

Patient: Thank you doctor. After doing all these I can upload all for further query.

Doctor: Hi, yes, upload in this query only. I will see and revert to you.

Patient: Thank you doctor. As per your guidelines, I have made one test. Due to city shutdown, I could not able to make an x-ray of chest. Fever is coming every 12 hours. But cough does not persist currently. The only problem is a fever. I am uploading the blood examination report. Please advice.

Doctor: Hi, I can understand your concern. I have gone through the report you have attached. (attachment removed to protect patient identity). Your total count is on the higher side of normal along with low hemoglobin. So start Azithromycin 500 mg 1-0-0, Tablet Dolo 650 1-0-1. Once your fever and cough subside, start an Iron tablet for low hemoglobin. I hope this helps.

Patient: Thank you doctor. How many tablets of Azithromycin to be taken? Is it in the morning? I want to make you know that I have already taken Azithromycin five days before. Should she start it? And also can I start an Iron tablet today itself?

Doctor: Hi, Yes, you can take iron tablet from today onwards. And no need for Azithromycin now as you have already taken it. Drink plenty of fluids orally and keep yourself hydrated. Do warm water gargles and steam inhalation four to five times a day.

Figure 1: An exemplar consultation in the English CovidDialog dataset. It consists of a brief description of the patient's medical conditions and the conversation between the patient and a doctor.

Description of medical conditions and history

疾病: 肺炎相关
(Disease: Pneumonia related)
病情描述: 咳嗽从 2 月开始到现在，有痰，怀疑是不是新冠状病毒？
(Description of medical condition: Cough from February to now. There is sputum. Am I infected by novel coronavirus?)
希望获得的帮助: 能不能医好这咳嗽病？
(Help needed: How to cure the cough?)
患病多久: 半年内
(Hong long the condition has been: Less than 6 months)
用药情况:
17：盐酸氨溴索片，头孢克肟片，肺力咳合剂；
21：盐酸氨溴索胶囊，头孢克肟片，喉疾灵胶囊，盐酸依匹斯汀胶囊；
23：盐酸依匹斯汀胶囊，盐酸克林霉素棕榈酸酯分散片，盐酸氨溴索口服溶液，喉疾灵胶囊；
26：肺力咳合剂，盐酸氨溴索胶囊，盐酸克林霉素棕榈酯分散片，复方甲氧那明胶囊，乙酰半胱氨酸颗粒；
(Medications:
17: Ambroxol hydrochloride tablets, Cefixime tablets, Feilike mixture;
21: Ambroxol hydrochloride capsules, Cefixime tablets, Houjiling capsules, and Epinastine hydrochloride capsules;
23: Epinastine hydrochloride capsules, Clindamycin hydrochloride palmitate dispersible tablets, Ambroxol hydrochloride oral solution, Houjiling capsules;
26: Feilike Mixture, Ambroxol Hydrochloride Capsules, Clindamycin Palmitate Dispersible Tablets, Compound Methionine Capsules, Acetylcysteine Granules;)
过敏史: 无
(Allergies: No)
既往病史: 无
(Past medical history: No)

Dialogue

医生: 感谢您的信任，病情资料我已详细阅读。是新冠活不了这么久的
(Doctor: Thank you for your trust. I have read the medical information in detail. If it was because of the coronavirus, you couldn't have lived that long.)
病人: 但是我吃了很多药都不见好，怎么办？
(Patient: But I took a lot of medications and it didn't work, what should I do?)
医生: 用点抗过敏药，比如西替利嗪
(Doctor: Use some antiallergic drugs, such as cetirizine.)
病人: 西替利嗪口服液？
(Patient: Cetirizine liquid?)
医生: 片
(Doctor: Tablets.)
病人: 我要不要去检查一下核酸？
(Patient: Should I get a RT-PCR test?)
医生: 不要浪费国家的资源了，不可能的
(Doctor: Impossible to be COVID. No need to do the test.)
病人: 我现在一动就会出很多汗，呼吸喉咙有点痒，咳痰。
(Patient: I sweat a lot when I move. The throat is itchy. There is sputum.)
医生: 你家几个人和你一样症状？
(Doctor: How many people in your family have the same symptoms as you?)
病人: 没有
(Patient: None)
……

Diagnosis and suggestions

病情摘要及初步印象: 过敏性咳嗽，焦虑抑郁状态
 (Summary of the condition and initial impressions: Allergic cough, anxiety and depression.)
总结建议: 西替利嗪片口服，必要时心理科就诊。
 (Summary of recommendations: Take cetirizine tablets orally, if necessary, see a psychologist.)

Figure 2: An exemplar consultation in the Chinese CovidDialog dataset. It consists of (1) description of medical conditions and history of the patient, (2) dialogue between doctor and patient, and (3) diagnosis and treatment suggestions given by the doctor.

sultations and 1232 utterances; (2) a Chinese dataset containing 399 consultations and 8440 utterances. The datasets are publicly available.

2. The English Dataset

The English CovidDialog dataset contains 603 consultations about COVID-19 and other related pneumonia, having 1232 utterances. The average, maximum, and minimum number of utterances in a conversation is 2.0, 17, and 2 respectively. The average, maximum, and minimum number of words in an utterance is 49.8, 339, and 2 respectively. Each consultation starts with a short description of the medical conditions of a patient, followed by the conversation between the patient and a doctor. The original dialogues are crawled from online healthcare forums, including icliniq.com[1], healthcaremagic.com[2], and healthtap.com[3].

3. The Chinese Dataset

The Chinese CovidDialog dataset contains 399 consultations about COVID-19 and other related pneumonia, having 8440 utterances. The average, maximum, and minimum number of utterances in a conversation is 15.5, 182, and 2 respectively. The average, maximum, and minimum number of Chinese characters in an utterance is 26.2, 405, and 4 respectively. Each consultation consists of three parts: (1) description of patient's medical condition and history; (2) conversation between patient and doctor; (3) (optional) diagnosis and treatment suggestions given by the doctor. In the description of patient's medical condition and history, the following fields are included: present disease, detailed description of present disease, what help is needed from the doctor, how long the disease has been, medications, allergies, and past disease. Figure ?? shows an exemplar consultation. The data is crawled from haodf.com[4], which is an online platform of healthcare services, including medical consultation, scheduling appointment with doctors, etc.

4. Conclusions

In response to the burnout of medical professionals and panic of the public, it is necessary to develop medical dialogue systems that are able to provide consultations to people for COVID-19 related topics. To facilitate the R&D of COVID-targeted dialogue systems, we build two COVID-related dialogue datasets in English and Chinese respectively. The datasets are open to the public.

---

1. https://www.icliniq.com/en_US/
2. https://www.healthcaremagic.com/
3. https://www.healthtap.com/
4. https://www.haodf.com/