Medical Question Understanding and Answering for Older Adults

Khalil Mrini¹, Chen Chen¹, Ndapa Nakashole¹, Nadir Weibel¹, Emilia Farcas²
¹Computer Science and Engineering, ²Qualcomm Institute
University of California San Diego
La Jolla, CA 92093
{khalil, chenchen, nnakashole, weibel, efarcas}@ucsd.edu

1 Introduction

Aging is one of the most significant social transformations in the 21st century [13] and enhancing the Quality of Life (QoL) of the aging population is crucial. In the US, around 49.2 million people are 65 or older, and by 2035, this number will increase to 78 million, projecting older adults to outnumber children for the first time [3]. Thus, there is a growing interest in technology catering to the well-being of older adults [15, 21] and their interactions with voice-based conversational agents [4, 20, 16]. For example, the EU has approved over €5M in 2018 for the NESTORE project, which aims to use body-attached mobile devices (e.g., mobile app and wearables) to monitor the health of older adults [6]. One identified key challenge is to use chatbot to understand simple user instructions effectively and efficiently [7].

Beyond this prior effort, we aim to design a voice-based conversational agent for older adults that promotes healthy aging, and provides understandable and explainable answers to health-related questions. In this paper, we demonstrate the progress of our ongoing project VOLI¹ (Voice Assistant for Quality of Life and Healthcare Improvement in Aging Populations) [10] and outline our future road maps to adapt to everyday speech used by older adults, and corresponding real-world user studies among communities of older adults.

2 Domain Adaptation for Everyday Speech

Studies find that one major challenge towards higher user retention of voice-based conversational agents for older adults is the rigidity of query commands [16]. How can we enable conversational agents to understand everyday speech to increase user retention among older adults?

Our prior work shows that the popular transformer architecture can leverage tree structures in formally written text through syntactic probing tasks, yet fail to do so in text written like everyday speech. Existing tree-structured transformers [14, 19] show state-of-the-art results on a wide range of tasks including sentiment analysis and text classification. However, their datasets are not user-written, but rather written in a formal, encyclopedia-like style. We introduce a novel tree-structured transformer architecture. Our method achieves state-of-the-art results in two widely used question answering (QA) benchmark datasets, but not in community question answering datasets, where text is user-written, long and informal. Through probing tasks, we show that not absorbing tree structure information leads to no increase in performance in QA. We hereby demonstrate a weakness in a popular NLU architecture to generalize to everyday speech.

For Natural Language Understanding (NLU) to generalize to user queries, we propose, in another work of our own, a question understanding approach that “translates” a user query into a formal question. Our approach is to augment datasets to cover both question summarization and recognizing

¹Project VOLI: http://voli.ucsd.edu

3rd Southern California NLP Symposium (SoCal NLP Symposium 2021), La Jolla, California (Online).
question entailment, and then to train fully shared parameters using a simple multi-task loss objective combining both tasks. We compare our results to single-task training on BART, the state-of-the-art in abstractive summarization [11]. We show across 4 medical datasets that our approach is efficient in low-resource settings and performs better than the BART baseline in at least one of human evaluation or ROUGE scores.

In a follow-up project, we propose a weighted multi-task loss objective, and a novel gradually soft parameter-sharing approach. In our novel approach, we train one fully shared encoder and two gradually soft-shared decoders. We add a loss term to constrain the 12 layers of the decoder to be close in representation space, such that the constraint is gradually loosened from the first layer until the last one, which is entirely task-specific. Our approach is a hybrid, encoder-decoder version of soft parameter-sharing [5] and the parameter-sharing configuration of MT-DNN [12], which has achieved state-of-the-art results on the popular GLUE benchmark [18]. We find through experiments that our approach outperforms other parameter-sharing configurations, as well as existing multi-task learning approaches using summarization.

Our next step is to ground question understanding models to a trusted medical knowledge base of answers. The US National Institutes of Health (NIH) released an FAQ-style dataset of medical question-answer pairs, called MedQuAD [1]. We also want to explore how pre-training on a knowledge base can help question understanding and answering systems. A large body of work has explored knowledge bases for question answering [24, 23, 25], and recently transformer models for knowledge bases have emerged [2].

We propose to train an end-to-end system that matches a user question to an FAQ and then to an expert-written answer. We first train the system to detect entailment between a user question and one of the FAQs in our dataset of medical question-answer pairs. Each FAQ is attached to multiple answers, and each answer has multiple sentences. In a second step, the model learns to select salient answer sentences given the user question. We propose to use pre-trained models from our prior work to create semi-supervised labels, and thus create a large medical answer sentence selection dataset. The user will then be provided with an expert-written answer, and can also compare their own question to the matched FAQ.

3 Real-World Evaluation by Older Adults

A practical inclusive technology design for older adults is important. We propose to integrate and deploy our framework among real-world aging individuals (aged 65 or older), recruited from the UC San Diego Outpatient Geriatric Primary Care Clinic.

The design and evaluations of existing voice-based conversational agents for enhancing healthcare management and quality of life only focused on general users, and relied on the training data collected primarily from younger adults [1]. In contrast, understanding voice interaction experience among aging individuals are challenging [17]. This is mainly caused by the degradation of their short-term memory, and results in barriers caused by their unique mental models [8]. In this work, we aim to bridge this gap by integrating our system with existing health care infrastructure, and deploying on the commercially available standalone smart speakers, placed in older adults’ residential houses (stand-alone home and retirement community).

As part of our preliminary work, we have conducted semi-structured interviews with 2 geriatricians, 3 nurses and 16 patients (aged 68 to 90), and identified the key barriers that older adults might encounter during daily life and while managing their health. Contextualized within ability-based design principles [22], we found that older adults and providers considered personalized chat, health advice and guidance toward heterogeneous everyday tasks as important needs.

To this end, based on our preliminary work and proposed framework, we aim to achieve real-world evaluations in terms of 2 folds. (1) A full-fledged evaluations of the effectiveness and efficiency of proposed framework. Qualitatively, we will evaluate the user experience by obtaining subjective comments, which can provide user feedback regarding fluency, coherence and comprehensibility of generated responses. Quantitatively, we will also capture the underlying metrics during the interactions between conversational agents and older adults, e.g., users’ response time and acknowledge intent regarding particular generated responses [9]. (2) An opportunity to contribute open-sourced and de-identified interaction dataset that would be potentially useful for researchers and practitioners.
References


