Medical Question Understanding and Answering for Older Adults

Khalil Mrini[1], **Chen Chen**[1], **Ndapa Nakashole**[1], **Nadir Weibel**[1], **Emilia Farcas**[2]
[1]Computer Science and Engineering, [2]Qualcomm 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 VOLI[1] (Voice Assistant for Quality of Life and Healthcare Improvement in Aging Populations) [10] and outline our future directions 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

---

[1]Project VOLI: `http://voli.ucsd.edu`

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

[1] Asma Ben Abacha and Dina Demner-Fushman. A question-entailment approach to question answering. *BMC bioinformatics*, 20(1):511, 2019.

[2] Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. Comet: Commonsense transformers for automatic knowledge graph construction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4762–4779, 2019.

[3] United States Census Bureau. Older people projected to outnumber children for first time in U.S. history. 03 2018.

[4] Kyungjin Chung, Young Hoon Oh, and Da Young Ju. Elderly users' interaction with conversational agent. In *Proceedings of the 7th International Conference on Human-Agent Interaction*, pages 277–279, 2019.

[5] Long Duong, Trevor Cohn, Steven Bird, and Paul Cook. Low resource dependency parsing: Cross-lingual parameter sharing in a neural network parser. In *Proceedings of the 53rd annual meeting of the Association for Computational Linguistics and the 7th international joint conference on natural language processing (volume 2: short papers)*, pages 845–850, 2015.

[6] Mira El Kamali, Leonardo Angelini, Maurizio Caon, Giuseppe Andreoni, Omar Abou Khaled, and Elena Mugellini. Towards the nestore e-coach: a tangible and embodied conversational agent for older adults. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, pages 1656–1663, 2018.

[7] Mira El Kamali, Leonardo Angelini, Maurizio Caon, Denis Lalanne, Omar Abou Khaled, and Elena Mugellini. An embodied and ubiquitous e-coach for accompanying older adults towards a better lifestyle. In *International Conference on Human-Computer Interaction*, pages 23–35. Springer, 2020.

[8] Arthur Fisk, Arthur D., Wendy Rogers, Wendy A., Neil Charness, Neil, Sara Czaja, John Sara, Sharit, and Joseph. *Designing for Older Adults: Principles and Creative Human Factors Approaches*. 01 2004.

[9] Seyyed Hadi Hashemi, Kyle Williams, Ahmed El Kholy, Imed Zitouni, and Paul A. Crook. Measuring user satisfaction on smart speaker intelligent assistants using intent sensitive query embeddings. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, CIKM '18, page 1183–1192, New York, NY, USA, 2018. Association for Computing Machinery.

[10] Janet G. Johnson, Khalil Mrini, Michael Hogarth, Alison A Moore, Ndapa Nakashole, Nadir Weibel, and Emilia Farcas. Voice-based conversational agents for older adults. In *Proceedings of the CHI 2020 Workshop on Conversational Agents for Health and Wellbeing*, Honolulu, Hawaii, 2020. Association for Computing Machinery.

[11] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, 2020.

[12] Xiaodong Liu, Pengcheng He, Weizhu Chen, and Jianfeng Gao. Multi-task deep neural networks for natural language understanding. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4487–4496, 2019.

[13] United Nations. World population ageing 2019 – highlights. 2020.

[14] Xuan-Phi Nguyen, Shafiq Joty, Steven Hoi, and Richard Socher. Tree-structured attention with hierarchical accumulation. In *International Conference on Learning Representations*, 2019.

[15] James Simpson, Franziska Gaiser, Miroslav Macík, and Timna Breßgott. Daisy: A friendly conversational agent for older adults. In *Proceedings of the 2nd Conference on Conversational User Interfaces*, pages 1–3, 2020.

[16] Milka Trajkova and Aqueasha Martin-Hammond. "alexa is a toy": Exploring older adults' reasons for using, limiting, and abandoning echo. CHI '20, page 1–13, New York, NY, USA, 2020. Association for Computing Machinery.

[17] John Vines, Gary Pritchard, Peter Wright, Patrick Olivier, and Katie Brittain. An age-old problem: Examining the discourses of ageing in hci and strategies for future research. *ACM Trans. Comput.-Hum. Interact.*, 22(1), February 2015.

[18] Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel Bowman. Glue: A multi-task benchmark and analysis platform for natural language understanding. In *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 353–355, 2018.

[19] Yaushian Wang, Hung-Yi Lee, and Yun-Nung Chen. Tree transformer: Integrating tree structures into self-attention. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1060–1070, 2019.

[20] Pierre Wargnier, Giovanni Carletti, Yann Laurent-Corniquet, Samuel Benveniste, Pierre Jouvelot, and Anne-Sophie Rigaud. Field evaluation with cognitively-impaired older adults of attention management in the embodied conversational agent louise. In *2016 IEEE International Conference on Serious Games and Applications for Health (SeGAH)*, pages 1–8. IEEE, 2016.

[21] Nirmalie Wiratunga, Kay Cooper, Anjana Wijekoon, Chamath Palihawadana, Vanessa Mendham, Ehud Reiter, and Kyle Martin. Fitchat: Conversational artificial intelligence interventions for encouraging physical activity in older adults. *arXiv preprint arXiv:2004.14067*, 2020.

[22] Jacob O. Wobbrock, Shaun K. Kane, Krzysztof Z. Gajos, Susumu Harada, and Jon Froehlich. Ability-based design: Concept, principles and examples. *ACM Trans. Access. Comput.*, 3(3), April 2011.

[23] Scott Wen-tau Yih, Ming-Wei Chang, Xiaodong He, and Jianfeng Gao. Semantic parsing via staged query graph generation: Question answering with knowledge base. 2015.

[24] Wen-tau Yih, Matthew Richardson, Christopher Meek, Ming-Wei Chang, and Jina Suh. The value of semantic parse labeling for knowledge base question answering. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 201–206, 2016.

[25] Mo Yu, Wenpeng Yin, Kazi Saidul Hasan, Cicero dos Santos, Bing Xiang, and Bowen Zhou. Improved neural relation detection for knowledge base question answering. *arXiv preprint arXiv:1704.06194*, 2017.