Trends and Overview: The Potential of Conversational Agents in Digital Health

Tulika Saha[1], Abhishek Tiwari[2], and Sriparna Saha[2]

[1] University of Liverpool, United Kingdom
sahatulika15@gmail.com
[2] Indian Institute of Technology Patna, India
abhisek_1921cs16@iitp.ac.in, sriparna.saha@gmail.com

1 Motivation

With the COVID-19 pandemic serving as a trigger, 2020 saw an unparalleled global expansion of tele-health [23]. Tele-health successfully lowers the need for in-person consultations and, thus, the danger of contracting a virus. While the COVID-19 pandemic sped up the adoption of virtual healthcare delivery in numerous nations, it also accelerated the creation of a wide range of other different technology-enabled systems and procedures for providing virtual healthcare to patients. Rightly so, the COVID-19 has brought many difficulties for patients[3] who need continuing care and monitoring for mental health issues and/or other chronic diseases.

One important technological advancement is the increasing use of Conversational Agents (CAs) or Virtual Assistants (VAs) in people's life, which now have numerous health applications. Chatbots or CAs communicate with users using a text-based or speech-based interface and consequently can make their services and applications available to a large segment of the population. Due in part with the on-set of COVID-19, CAs are being used more frequently in the healthcare industry as a promising tool to enhance delivery and quality. Numerous healthcare surveys conducted over the past years have revealed a worrying shortage of doctors [14] compared to the doctor-to-population ratio in physical health and even more severe for mental health. Thus, CAs in healthcare is becoming more and more popular, driven by the need to assist the doctors and utilize their time effectively. The usage of CAs by patients and medical personnel has also shown to be well accepted, with high ratings for perceived utility, convenience, and participation in overcoming service and logistical constraints. The recent advancement in messaging services amongst leading social media firms and the latest rally to develop automated systems has driven novel research in this area. Thus, it has become imperative, more than ever to focus on understanding and analysing growing trends of human-computer interfaces, i.e., VAs which will further pave way for developing robust computational models in healthcare including mental health. The motivation behind this tutorial is to analyze the growing trend of

[3] https://www.who.int/news/item/02-03-2022-covid-19-pandemic-triggers-25-increase-in-prevalence-of-anxiety-and-depression-worldwide

VAs in healthcare and provide the IR researchers with an overall perspective of where the AI and NLP communities are heading which can further pave way for ground-breaking novelties benefitting the research community and the society at large.

2 Objective

By attending this tutorial, participants will learn about: **(i)** the existing limitations in digital health in the realms of NLP specifically VAs and the recent upcoming frameworks which alleviate these limitations and enhance their capabilities; **(ii)** the basics of a Dialogue System aka Conversational AI and the recent most emerging branches of communication with the end user; **(iii)** the most successful techniques addressing multiple sub-modules of a VA for both physical and mental health; **(iv)** for practical exposure, the tutorial will also provide a live demonstration of the recently published *symptom investigation and disease diagnosis VA* and **(v)** some pressing challenges to the adoption of AI in healthcare, such as Reliability, Explainability, and Safety (RES) as well as discuss possible remedies to accelerate progress in this area.

3 Relevance to the IR Community

In the early 20s, research in digital health specifically VAs was unheard of, largely because of the communities disbelief of whether an automated system can be accurate, knowledgeable, intelligent and human enough to make decisions in the sensitive domain of health. With the advent of Deep Learning and massive expansion of e-commerce in the mid 20s through recommender systems, CAs etc. paved the way for novel ideas and digitisation of health goals.

Initially, the research in digital or tele-health were widely focused on social media analysis or support forums due to the lack of real-time data available owing to privacy concerns. Soon enough the NLP conferences were flooded with tele-health oriented papers related to social media. However, in the recent times there has been a massive shift towards real-time data processing and inference where researchers realized that the overall health of the society can be benefitted while only dealing with real-time data between patients and doctors. During this shift, the IR community remained consumers but not key innovators in healthcare VAs, largely falling in the track of AI for Social Good.

It is important for the IR community to resume greater focus on digital health, guided by the need of CAs in healthcare including mental health. In the context of IR, there are two main threads of work in an end-to-end CA: conversational Q&A and conversational recommendation. Currently, these are viewed as two distinct systems with unique objectives, architectures, and evaluation standards. Instead, the two should be seamlessly integrated into CAs to better help users, changing the perspective from one that is isolated to one that is more unified. In order to actively encourage good engagement, the multi-modality of interactions also needs to be more thoroughly acknowledged.

4 Tutorial logistics

The tutorial will interleave slide-based presentation, scribbling on a whiteboard, screen-sharing short demos, and Q&A sessions (at least every 25 minutes). The tutorial organisers may set up Slack or Piazza for attendees to interact with them and with each other. The schedule is for 165 minutes of presentation and 15 minutes of Q&A suitable for introductory to advanced target audience.

4.1 Preliminaries (20 minutes)

• **Dialogue System.** The tutorial will introduce the basics of a typical Dialogue System (DS). A DS is primarily known to comprise of three prime modules : (i) Natural Language Understanding; (ii) Dialogue Policy Learning also known as Dialogue Management (DM) Strategy and (iii) Natural Language Generation also known as Response Generation in dialogues. Lately, two branch of DSs have emerged focused on varying ways of communicating with the user : (i) Modularized DS and (ii) End-to-end Generation Framework. Each of these variation and its workflow will be inspected in details.
• **Need and Utility of VAs in Healthcare.** The range of CAs in healthcare has been discussed in a number of recent review studies. A significant majority of CAs in healthcare have been categorised as goal-oriented agents since they have been built to assist patients and healthcare workers in certain tasks. For CAs in health, a taxonomy has been created in recent times which categorizes six health-related purposes into training, education, aid, prevention, diagnostic, and assistance for the elderly. The tutorial will cover in details the different use cases concerning the above mentioned six purposes which best describes the role of VAs in healthcare.
• **Challenges in Healthcare VAs.** Although AI has been a tremendous success for healthcare in recent years, there are a few pressing challenges for healthcare VAs which needs to be addressed in the upcoming years. For e.g., adequate amount of data due to privacy concerns, risk associated with the failure of AI models, explainable decision support system etc. will be discussed in details.

4.2 Natural Language Understanding (45 minutes)

Natural Language Understanding (NLU) module is responsible for making sense of the user input. It extracts various information from the user utterance (coarse to fine-grained) which is used further in the conversational framework. NLU module is typically framed as a classification task. Formally, consider a training set, $\{x_i, y_i\}_{i=1}^{N}$, where x_i is the i-th input sample, y_i is the label vector for the classification task, respectively and N is the number of training instances. $f(x, w)$ denotes the classifier and w is its parameters. The task is to find the optimal parameter, w^*, by minimizing the training loss, $1/N \sum_{i=1}^{N} L^{train}(w)$, where $L^{train}(w) = l(y_i, f(x_i, w))$. In healthcare VAs, different type of NLU modules have been developed in the recent times covering both physical and mental health ranging from multi-modality to multi-tasking frameworks.

• **Physical Health.** Lately, disease diagnosis VA is an upcoming research area. Understanding patients' concerns from their utterances are critical to diagnosis and treatment outcomes [4]. The recent works on medical dialogue understanding can broadly be grouped into two categories: (a) pre-trained transformer-based joint intent and entity model and (b) Multi-label entity classification [20]. The community is also slowly advancing towards multi-modal signal processing in digital health. Primarily, when we consult with doctors, we often show our signs/symptoms through visuals. Thus, an symptom image identification is an integral part of medical disease diagnosis DS. Authors of [21] introduce a context aware image identification model, which incorporates conversation history for identifying an image adequately.
• **Mental Health.** Mental health dialogue understanding is somewhat an implicit task as opposed to disease diagnosis VAs [21] where users are expected to provide task information explicitly. Thus, these mental health VAs [16] also utilize semantic features such as sentiment and emotion for understanding users concern accurately and to serve more appropriately. Lately, researchers are focused on identifying mental health disorder [17], symptom investigation [12], gender prediction [11] etc. from conversations either collected from support forums [5] or from real counselling settings [1]. Multi-modality has also been explored in the recent times to identify mental health diseases such as depression [13].

4.3 Dialogue Policy Learning (40 minutes)

Dialogue Policy Learning also known as DM strategy is responsible for deciding the flow of conversation in any DS. It takes as input the information extracted by the NLU module in its state space and outputs an appropriate action based on a policy to communicate with the user so as to maximize a reward (short or long term goals). DM is often viewed as a sequential decision problem and is formulated as a Markov Decision Process (MDP) which optimizes the dialogue policy through a Reinforcement Learning (RL) algorithm. Formally, MDP can be represented as a five-tuple (S, A, P, R, π), where S is the dialogue state, A denotes the set of possible actions for the VA, P signifies transition probability $(P(S_{t+1}|S_t))$ and R denotes reward model.
• **Physical Health.** For a large number of non-fatal diseases, doctors typically identify patients' diseases by conducting a symptom examination through conversations only. Inspired by such real-world scenarios, the researchers have formulated automatic disease diagnosis as a task-oriented dialogue framework [22]. Authors of [24] further proposed a knowledge routed relational dialogue system that incorporates an external rich medical knowledge graph into dialogue policy learning for knowledge grounded topic transition. To overcome the rule-based dependency of medical department identification, Liao et al. [8] proposed a novel policy learning framework for symptom investigation and disease diagnosis using Hierarchical Reinforcement Learning (HRL). The community is also advancing towards multi-modal VAs to provide end-users with a human-like experience. Motivated by the importance of visual form of symptom reporting, researchers proposed a HRL based multi-modal disease diagnosis VA [21].

• **Mental Health.** When talking to someone, our physical presence is usually appreciated, and it becomes more crucial if it is related to a serious concern like anxiety or depression. In [3], the authors have developed an animated virtual interviewer SimSensei Kiosk to create engaging face-to-face interactions with mental health support seekers.

4.4 Generation Frameworks (45 minutes)

The generation framework typically skips the DM module and is focused on generating the next response of the VA aligned with the user context. Formally, given a user utterance, $X_t = (x_{t,1}, x_{t,2}, ..., x_{t,n})$, a conversational context/history, $C = (c_1, c_2, ..., c_{t-1})$, where $c_i = (X_i, Z_i)$, the task is to generate next textual response of the VA, $Z_t = (z_{t,1}, z_{t,2}, ..., z_{t,n''})$. In the literature, three popular approaches have been employed for generating VAs response: (a) template-based response modeling [24] (b) Seq2Seq [7] and (c) pre-trained generation frameworks [2]. *The tutorial will also demonstrate Seq2Seq and fine-tuning of GPT models for dialogue response generation task.*

• **Physical Health.** In real-time, doctors' investigation also depends on patients' personal information, such as age and gender, in addition to patients' reported major difficulties. Inspired by such scenarios, the authors have proposed a context-aware HRL-based dialogue system [6] for symptom investigation followed by disease prediction. Authors of [9] have proposed a graph-based dialogue generation framework that utilizes commonsense knowledge for identifying new diseases. In the diagnostic process, external medical knowledge aids clinicians in narrowing their investigation space and efficiently utilize the gathered information. Motivated by the observation, the authors [10] have proposed medical knowledge graph guided response generation model.

• **Mental Health.** Extensive research has been carried out in mental health addressing the aspects of empathy and motivation which have been identified as key affective factors providing positive outcome in support based conversations. In [19], authors presented a computational approach to understand empathy based on three communication mechanisms : *Emotional Reactions*, *Explorations* and *Interpretations*. Furthermore, in [18], authors developed an empathetic rewriting framework, named *PARTNER* that transforms lower empathetic responses into higher empathetic content. Authors of [16] proposed a VA acting as the first point of contact for mentally distressed support seekers afflicted with some form of mental illness. Authors of [15] went ahead and combined both these aspects of empathy and motivation in a unified end-to-end system for online mental health support.

4.5 Overall Summary & Scope for Future Work (15 minutes)

CAs are an emerging technology for use in healthcare that has not yet undergone a thorough evaluation. Future studies should concentrate on evaluating the viability, acceptability, safety, and efficacy of various CA forms that are in line with the requirements and preferences of the target audience. Additionally, there is a need for more in-depth research on the function of CAs in the current health systems as well as clearer guidelines for the creation and evaluation of CAs connected to healthcare.

5 Supporting Material

The tutorial will provide the following material: (1) lecture video recording, (2) annotated slides, (3) assignments, exams and projects from related courses taught, (4) extended bibliography, and (5) compendium of public software and data sets.

6 Presenter information

• **Dr. Tulika Saha** (corresponding author) is a Lecturer of Computer Science at the University of Liverpool, United Kingdom (UK). Her current research interests include ML, DL, NLP typically Dialogue Systems, AI for Social Good, Social Media Analysis etc. She was a postdoctoral research fellow at the National Centre for Text Mining, University of Manchester, UK. Previously she earned her Ph.D. from Indian Institute of Technology Patna, India. Her research articles are published in top-tier conferences such as ACL, ACM SIGIR etc. and peer-reviewed journals.
• **Abhisek Tiwari**is a research scholar (Prime Minister Research Fellow) in Computer science and Engineering, Indian Institute of Technology, Patna. His research interest includes AI for Social Good, NLP, typically Conversational AI, and RL. He is also serving as a guest lecturer at NSIT Bihta, India. His research works have been published in reputable conferences, such as CIKM, IJCNLP, and peer-reviewed journals. Abhisek has delivered several tutorials including the GIAN Course on DL Techniques for Conversational AI, conducted birds-of-a-feather sessions at top-tier conferences such as ACL, ICLR, and NeurIPS.
• **Dr. Sriparna Saha** is currently serving as an Associate Professor (h5-index:33, total citations: 6201 as per Google Scholar), Head of Department in Computer Science and Engineering, Indian Institute of Technology Patna, India (https://www.iitp.ac.in/∼sriparna/). Her current research interests include ML, DL, NLP, AI for Social Good, Information Retrieval. She has published more than 400 papers in reputed journals and conferences including ACL, SIGIR, AAAI, EMNLP, ECIR, COLING, ACM MM etc. Her tutorial on "Summarization Systems: From Text to Multimodal" is accepted to be delivered in ICONIP 2022 to be held in New Delhi, India. She is one of the special session organizers of ICONIP 2021 on the topic of "Smart Home Technologies & Services for the Wellbeing and Sustainability of Society". She was one of the special session organizers of IEEE SSCI 2021 on the topic of "Computational Intelligence for Natural Language Processing". She has delivered a tutorial session on "Multimodality Helps in Solving Biomedical Problems: Theory and Applications" in IEEE WCCI 2020.

References

1. Althoff, T., Clark, K., Leskovec, J.: Large-scale analysis of counseling conversations: An application of natural language processing to mental health. Transactions of the Association for Computational Linguistics **4**, 463–476 (2016)
2. Chen, X., Meng, F., Li, P., Chen, F., Xu, S., Xu, B., Zhou, J.: Bridging the gap between prior and posterior knowledge selection for knowledge-grounded dialogue generation. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 3426–3437 (2020)
3. DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., Georgila, K., Gratch, J., Hartholt, A., Lhommet, M., et al.: Simsensei kiosk: A virtual human interviewer for healthcare decision support. In: Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems. pp. 1061–1068 (2014)
4. Enarvi, S., Amoia, M., Teba, M.D.A., Delaney, B., Diehl, F., Hahn, S., Harris, K., McGrath, L., Pan, Y., Pinto, J., et al.: Generating medical reports from patient-doctor conversations using sequence-to-sequence models. In: Proceedings of the first workshop on natural language processing for medical conversations. pp. 22–30 (2020)
5. Ji, S., Li, X., Huang, Z., Cambria, E.: Suicidal ideation and mental disorder detection with attentive relation networks. Neural Computing and Applications **34**(13), 10309–10319 (2022)
6. Kao, H.C., Tang, K.F., Chang, E.: Context-aware symptom checking for disease diagnosis using hierarchical reinforcement learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 32 (2018)
7. Li, D., Ren, Z., Ren, P., Chen, Z., Fan, M., Ma, J., de Rijke, M.: Semi-supervised variational reasoning for medical dialogue generation. In: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 544–554 (2021)
8. Liao, K., Liu, Q., Wei, Z., Peng, B., Chen, Q., Sun, W., Huang, X.: Task-oriented dialogue system for automatic disease diagnosis via hierarchical reinforcement learning. arXiv preprint arXiv:2004.14254 (2020)
9. Lin, S., Zhou, P., Liang, X., Tang, J., Zhao, R., Chen, Z., Lin, L.: Graph-evolving meta-learning for low-resource medical dialogue generation. In: Proceedings of the 35th AAAI Conference on Artificial Intelligence. pp. 13362–13370. AAAI Press (2021)
10. Liu, W., Tang, J., Liang, X., Cai, Q.: Heterogeneous graph reasoning for knowledge-grounded medical dialogue system. Neurocomputing **442**, 260–268 (2021)
11. Lokala, U., Srivastava, A., Dastidar, T.G., Chakraborty, T., Akhtar, M.S., Panahiazar, M., Sheth, A.: A computational approach to understand mental health from reddit: Knowledge-aware multitask learning framework. In: Proceedings of the International AAAI Conference on Web and Social Media. vol. 16, pp. 640–650 (2022)
12. Patra, B.G., Kar, R., Roberts, K., Wu, H.: Mental health severity detection from psychological forum data using domain-specific unlabelled data. AMIA Summits on Translational Science Proceedings **2020**, 487 (2020)
13. Qureshi, S.A., Saha, S., Hasanuzzaman, M., Dias, G.: Multitask representation learning for multimodal estimation of depression level. IEEE Intelligent Systems **34**(5), 45–52 (2019)
14. Rasmussen, K., Belisario, J.M., Wark, P.A., Molina, J.A., Loong, S.L., Cotic, Z., Papachristou, N., Riboli-Sasco, E., Car, L.T., Musulanov, E.M., et al.: Offline

elearning for undergraduates in health professions: a systematic review of the impact on knowledge, skills, attitudes and satisfaction. Journal of global health **4**(1) (2014)
15. Saha, T., Gakhreja, V., Das, A.S., Chakraborty, S., Saha, S.: Towards motivational and empathetic response generation in online mental health support. In: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 2650–2656 (2022)
16. Saha, T., Reddy, S., Das, A., Saha, S., Bhattacharyya, P.: A shoulder to cry on: Towards a motivational virtual assistant for assuaging mental agony. In: Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. pp. 2436–2449 (2022)
17. Saha, T., Reddy, S.M., Saha, S., Bhattacharyya, P.: Mental health disorder identification from motivational conversations. IEEE Transactions on Computational Social Systems (2022)
18. Sharma, A., Lin, I.W., Miner, A.S., Atkins, D.C., Althoff, T.: Towards facilitating empathic conversations in online mental health support: A reinforcement learning approach. In: Leskovec, J., Grobelnik, M., Najork, M., Tang, J., Zia, L. (eds.) WWW '21: The Web Conference 2021, Virtual Event / Ljubljana, Slovenia, April 19-23, 2021. pp. 194–205. ACM / IW3C2 (2021). https://doi.org/10.1145/3442381.3450097, https://doi.org/10.1145/3442381.3450097
19. Sharma, A., Miner, A.S., Atkins, D.C., Althoff, T.: A computational approach to understanding empathy expressed in text-based mental health support. In: Webber, B., Cohn, T., He, Y., Liu, Y. (eds.) Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020. pp. 5263–5276. Association for Computational Linguistics (2020). https://doi.org/10.18653/v1/2020.emnlp-main.425, https://doi.org/10.18653/v1/2020.emnlp-main.425
20. Shi, X., Hu, H., Che, W., Sun, Z., Liu, T., Huang, J.: Understanding medical conversations with scattered keyword attention and weak supervision from responses. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 8838–8845 (2020)
21. Tiwari, A., Manthena, M., Saha, S., Bhattacharyya, P., Dhar, M., Tiwari, S.: Dr. can see: Towards a multi-modal disease diagnosis virtual assistant. In: Proceedings of the 31st ACM International Conference on Information & Knowledge Management. pp. 1935–1944 (2022)
22. Wei, Z., Liu, Q., Peng, B., Tou, H., Chen, T., Huang, X.J., Wong, K.F., Dai, X.: Task-oriented dialogue system for automatic diagnosis. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). pp. 201–207 (2018)
23. Wosik, J., Fudim, M., Cameron, B., Gellad, Z.F., Cho, A., Phinney, D., Curtis, S., Roman, M., Poon, E.G., Ferranti, J., et al.: Telehealth transformation: Covid-19 and the rise of virtual care. Journal of the American Medical Informatics Association **27**(6), 957–962 (2020)
24. Xu, L., Zhou, Q., Gong, K., Liang, X., Tang, J., Lin, L.: End-to-end knowledge-routed relational dialogue system for automatic diagnosis. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 7346–7353 (2019)