Doctoral Student in Computer Science, with Specialization in Artificial Intelligence

Ref no: ORU 2.1.1-08267/2022

Research Proposal

Applicant: Konark Karna, MSc Advanced Computer Science, Northumbria University

1. Aim

Advancing human-machine communication, where agent proactively take lead into conversation, and adapt to subject evolving emotional state/response

2. Background

As robots become more autonomous and increasingly integrated into human lives, it is essential to establish effective communication between humans and machines for personalized, efficient interaction or to achieve common goals, in the case of multi-agent systems.

Human-Machine Communication (HMC) research, therefore, must provide solutions to how machines can understand and interpret human communication, speech, gestures, expression, and body languages, and thereby respond in a human-like manner. Recently, there has been a great interest among researchers to improve HMC, and many aspects of HMC have been explored, paving the way for future research.

In the case of a single machine in interaction, HMC is classified into two types: proactive conversation, and knowledge grounded conversation. Proactive conversation endows the system to lead conversation, and it is helpful in task-oriented systems such as booking a table in a restaurant. Knowledge grounded conversation, on the other hand, requires communication to be extracted from the structured or unstructured knowledge graph in the system at each time step, to have informative exchange between human and machine (Wu et al., 2019).

In cases of multi-agent systems, communication is classified into discrete and continuous communication. In continuous, agents communicate via a connecting vector which gives each agent asses into another's internal state, whereas in discrete communication it is inaccessible (Lazaridou and Baroni, 2020). They also surveyed degrees to measure effective communication on positive signalling (consists of context independence and speaker consistency), and positive listening (consists of instantaneous coordination, and casual influence of communication). Other ways are topographic similarity and compositionality (Lazaridou and Baroni, 2020).

For single machine-human interaction, many datasets have been proposed to train newer HMC methods. For proactive conversation, a proposed Action-Based Conversation Dataset (ABCD) of 10K human-to-human interaction with 55 user intent, is showing impressive

results for an in-depth task-oriented dialogue system. (Chen et al., 2021). It discussed two new dialogue tasks, Action State Tracking (AST) to adjudge customer intent, and Cascading Dialogue State (CDS) giving agent to respond with additional options and success to be measured over the entire sequence of communication. For AST, BERT is used for context input, and eventually TRADE (Wu et al. 2019) architecture is used to get prediction, and then CDS again uses an intuitive method of using BERT followed by recent proposed ranker design (Guu et al. 2020).

Similarly, DuConv dataset is proposed with 30K dialogues (Wu et al., 2019). This dataset has four steps i.e., knowledge crawling, knowledge graph construction, conversation goal assignment and conversation crowdsourcing. Afterwards, researchers have proposed two methods, retrieval-based method and generation-based method to store knowledge and select appropriate response as per user inputs. Both methods are intuitive development on the Transformer-based method.

In addition to aforesaid transformer-based recently developed methods, researchers have also shown great progress with Deep Reinforcement Learning (DRL) in HMC. (Chen et al., 2018) proposed multi-agent dialogue policy (MADP) uses S-Agent and G-Agent to reach faster policy learning for dialogue exchange, which can be learnt through any DRL algorithm. S-Agent in MADP follows continuous and discrete communication as discussed earlier with shared and private parameters, respectively. We also have AgentGraph methods proposed by the same group of researchers which combines Graph Neural Network, with DRL methods to further improve on faster convergence (Chen et al., 2019).

Furthermore, we again have research into gesture and speech recognition in HMC. (Mazhar et al., 2019) created their own dataset openSign and using YOLO-2 with Inception V3 as reference network, later detected with 98% accuracy on hand-gesture detection in real-time. Similarly, (Gao et al., 2020) used SSD with ResNet-101 as a reference network to robust real-time hand gesture detection in futuristic space human-robot interaction scenarios. Next, we have research discussing Dialogue Affective (DA) where machines could recognise human emotion at the end of their speech, and if emotion prediction is high, machine gets to respond correspondingly (Li et al., 2023). In this paper, researchers have also proposed a laughter prediction where machines with Bi-Direction GRU could detect laughter probability to affective dialogue with users.

These developments in HMC are inspiration for many new researchers to start exploring ahead on these methods intuitively to develop more and more affective and proactive human machine interaction methods at the earliest, considering all the ethical issues that can get into such research work such as in the collection of larger datasets.

3. Methodology

First of all, an absolute thorough literature review is needed to build complete knowledge on Human-machine interaction.

Afterwards, a development of more comprehensive dataset can be a good start, especially for proactive HMC. Many new methods can also be worked upon to further improve on Transformers or BERT based dialogue generation from the decoder side of our HMC model. In cases of emotion recognition of human, we still need research to recognize more emotion

and effectively such as frustration, sadness, excitement, confusion from the text, and wishfully from the facial expression recognition using faster SSD detection models, for instance. We again would need to develop a dataset for such tasks. This dataset and model framework must also be standard to get integrated into physical human robot interaction library, OpenPHRI.

I am personally excited to see DRL methods being incorporated into these research, and more evolved DRL methods such as Trust Region Policy Optimization (TRPO), PPO, etc can be evaluated as well for their effectiveness in dialogue exchange between human and robots, especially in the cases of multi-agent systems, and proactive communication. Similarly, Graph Neural Network and its variants are going to play an instrumental role in further development of knowledge grounded conversation.

Furthermore, we have a few researchers using contrastive learning into HMC for facial expression recognition and faster hand-gestures recognition. Contrastive Learning with framework SimCLR is the most promising of self-supervised learning and is an ingenious method in AI when we need our models to learn labels. Adopting contrastive learning can be absolutely effective to learn multiple facial expression, especially when some portion of face remain hidden. Thus, providing clearer understanding of user's emotional state and decide agents' response thereafter.

However, with more extensive literature, many new ideas can germinate to further explore into AI methods for improved HMC, while simultaneously developing deeper understanding of other proposed methods that, perhaps, can be intuitively tweaked as well to improve HMC. Lastly, strong background in mathematics will be crucial to develop any newer method.

4. References

Chen, D., Chen, H., Yang, Y., Lin, A., Yu, Z., 2021. Action-Based Conversations Dataset: A Corpus for Building More In-Depth Task-Oriented Dialogue Systems. *arXiv*. Available at: http://arxiv.org/abs/2104.00783

Chen, L., Chang, C., Chen, Z., Tan, B., Gasic, M., Yu, K., 2018. Policy Adaptation for Deep Reinforcement Learning-Based Dialogue Management. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). ICASSP 2018 - 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, AB: IEEE, pp. 6074–6078. Available at: https://doi.org/10.1109/ICASSP.2018.8462272.

Chen, L., Chen, Z., Tan, B., Long, S., Gasic, M., Yu, K., 2019. AgentGraph: Towards Universal Dialogue Management with Structured Deep Reinforcement Learning. *arXiv*. Available at: http://arxiv.org/abs/1905.11259

Gao, Q., Liu, J., Ju, Z., 2020. Robust real-time hand detection and localization for space human–robot interaction based on deep learning. *Neurocomputing, 390*, pp. 198–206. Available at: https://doi.org/10.1016/j.neucom.2019.02.066

Guu, K., Lee, K., Tung, Z., Pasupat, P. and Chang, M., 2020, November. Retrieval augmented language model pre-training. In *International conference on machine learning* (pp. 3929-3938). PMLR.

Mazhar, O., Navarro, B., Ramdani, S., Passama, R., Cherubini, A., 2019. A real-time human-robot interaction framework with robust background invariant hand gesture detection. *Robotics and Computer-Integrated Manufacturing 60*, 34–48. Available at: https://doi.org/10.1016/j.rcim.2019.05.008

Lazaridou, A., Baroni, M., 2020. Emergent Multi-Agent Communication in the Deep Learning Era. *arXiv.* Available at: http://arxiv.org/abs/2006.02419

Li, Y., Inoue, K., Tian, L., Fu, C., Ishi, C., Ishiguro, H., Kawahara, T., Lai, C., 2023. I Know Your Feelings Before You Do: Predicting Future Affective Reactions in Human-Computer Dialogue. *arXiv.* Available at: http://arxiv.org/abs/2303.00146

Wu, W., Guo, Z., Zhou, X., Wu, H., Zhang, X., Lian, R., Wang, H., 2019. Proactive human-machine conversation with explicit conversation goal. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Available at: https://doi.org/10.18653/v1/p19-1369.

Wu, C.S., Madotto, A., Hosseini-Asl, E., Xiong, C., Socher, R. and Fung, P., 2019. Transferable multi-domain state generator for task-oriented dialogue systems. *arXiv preprint.* Avaiable at: https://arxiv.org/abs/1905.08743