

# History-Aware Hierarchical Transformer for Multi-session Open-domain Dialogue System

Tong Zhang<sup>1</sup>, Yong Liu<sup>2,3</sup>, Boyang Li<sup>1</sup>, Zhiwei Zeng<sup>3</sup>, Pengwei Wang<sup>4</sup>, Yuan You<sup>4</sup>  
Chunyan Miao<sup>1,2,3</sup>, Lizhen Cui<sup>5</sup>

<sup>1</sup>School of Computer Science and Engineering, Nanyang Technological University, Singapore

<sup>2</sup>Alibaba-NTU Singapore Joint Research Institute, Nanyang Technological University, Singapore

<sup>3</sup>Joint NTU-UBC LILY Research Centre, Nanyang Technological University, Singapore

<sup>4</sup>Alibaba Group, China

<sup>5</sup>School of Software, Shandong University, China

## Abstract

With the evolution of pre-trained language models, current open-domain dialogue systems have achieved great progress in conducting one-session conversations. In contrast, Multi-Session Conversation (MSC), which consists of multiple sessions over a long term with the same user, is under-investigated. In this paper, we propose History-Aware Hierarchical Transformer (HAHT) for multi-session open-domain dialogue. HAHT maintains a long-term memory of history conversations and utilizes history information to understand current conversation context and generate well-informed and context-relevant responses. Specifically, HAHT first encodes history conversation sessions hierarchically into a history memory. Then, HAHT leverages historical information to facilitate the understanding of the current conversation context by encoding the history memory together with the current context with attention-based mechanisms. Finally, to explicitly utilize historical information, HAHT uses a history-aware response generator that switches between a generic vocabulary and a history-aware vocabulary. Experimental results on a large-scale MSC dataset suggest that the proposed HAHT model consistently outperforms baseline models. Human evaluation results support that HAHT generates more human-like, context-relevant and history-relevant responses than baseline models.

## 1 Introduction

Open-domain dialogue systems, also known as chatbots, are designed to converse with and engage users on any topic with the aim of establishing, maintaining, and strengthening long-term relationships (Clark et al., 2019; Roller et al., 2020). Recently, open-domain dialogue systems built based on large-scale generative pre-trained models (Adiwardana et al., 2020; Roller et al., 2021; Zhang et al., 2020) have substantially improved the performance of chatbots.

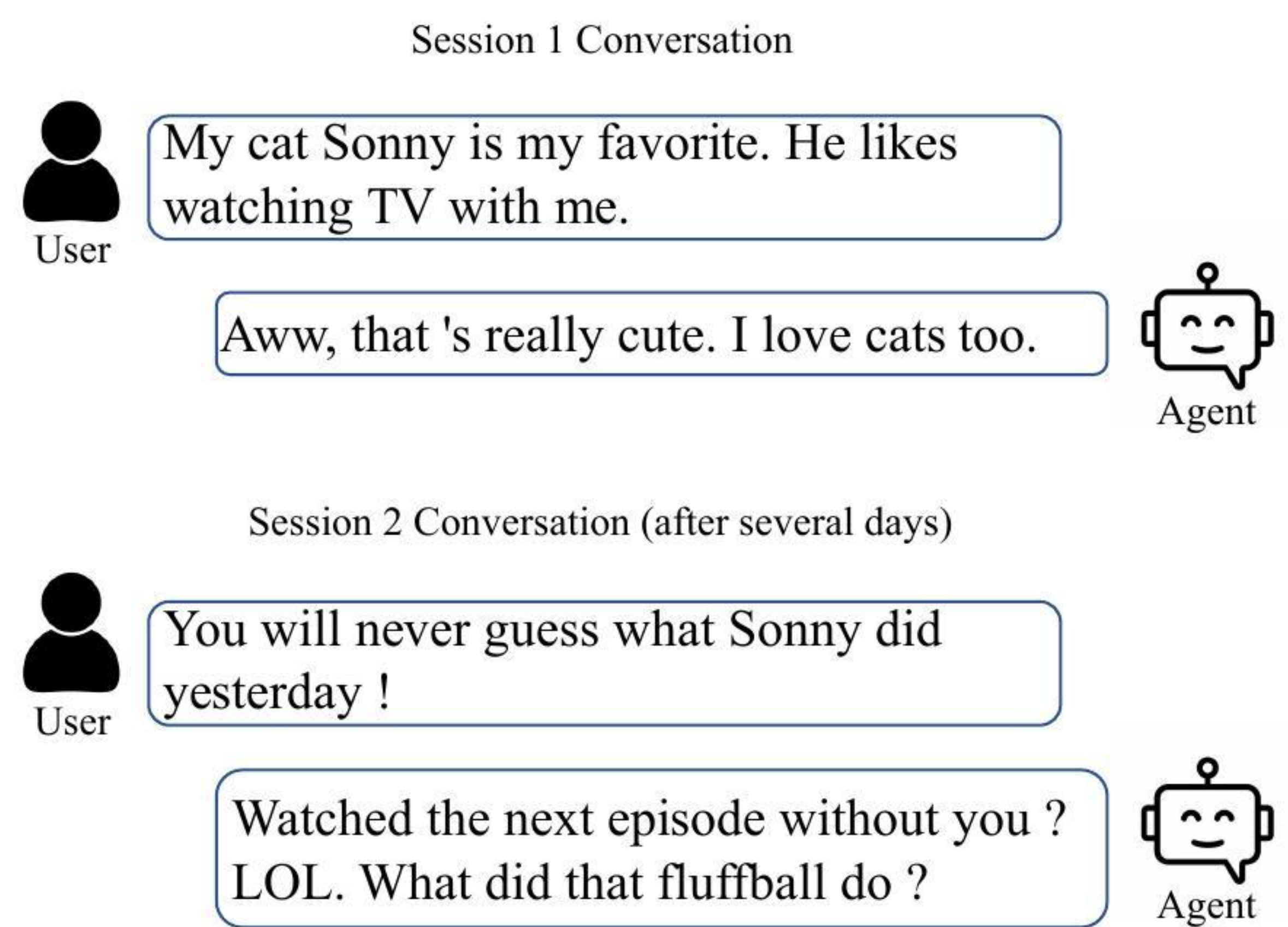


Figure 1: An illustrated example of a two-session conversation between a user and an agent.

However, most existing chatbots are designed to interact with users in a single conversation session. When the current session ends, the chatbot forgets its contents and will commence a new independent session with the same user next time. When previously discussed topics reemerge, such chatbots often appear ignorant and fail to reengage users appropriately. The apparent forgetfulness limits the chatbots' ability to establish and maintain long-term relationships with users.

We argue that, to better engage users in multi-session conversations (MSCs), a chatbot should maintain a long-term memory of historical contexts, which allows the chatbot to reengage the user appropriately when similar contexts reemerge. By learning from historical conversations, the chatbot should gradually refine its understanding of and deepen its relationship with the user. Figure 1 shows an example of a two-session conversation between a user and a chatbot. In the second session, the chatbot infers that Sonny is a cat and generates the response based on the history information that Sonny likes watching TV with the user.

History-aware chatbots will be able to generate more well-informed and context-relevant re-