Prompt Compression for Dialog Navigation with Need-Oriented Environmental Knowledge Base and Large Language Models

Hiroaki Shimoma[1,*,†], Sudesna Chakraborty[1,†] and Takeshi Morita[1,2,†]

[1]*Aoyama Gakuin University, Kanagawa, Japan*

[2]*National Institute of Advanced Industrial Science and Technology, Kotoku, Japan*

Abstract

In Embodied AI, navigation agents using Large Language Models (LLMs) often rely on lengthy prompts that include extensive environmental information. This becomes increasingly problematic in complex environments such as the VirtualHome simulator, where incorporating all object data can reduce accuracy and increase computational costs. To address this, we propose a prompt compression technique based on a "need-oriented" environmental knowledge base. Our system first infers a user's underlying need from their natural language request using Murray's theory of human needs. It then retrieves only the objects relevant to that need from our knowledge base. A compressed prompt, containing only the user's request and the specific objects, is then sent to the LLM. The results showed this method significantly improves navigation accuracy while reducing prompt length compared to approaches that use all environmental data.

Keywords

need-oriented environmental knowledge base, prompt compression, navigation, large language models

1. Introduction

In Embodied AI [1], using Large Language Models (LLMs) for dialog-based navigation [2] is a promising approach. However, a key challenge is that embedding all environmental knowledge into prompts makes them lengthy, increasing computational costs and reducing accuracy.

To address this challenge, our study proposes a prompt compression technique for navigation system in the VirtualHome (VH) simulator [3]. The method uses a novel "need-oriented" environmental knowledge base, inspired by Murray's theory of human needs, to select only the essential information required for the LLM's decision-making. This approach aims to improve inference performance and reduce operational costs, creating more efficient and scalable LLM-based household agents.

2. Related Work

While earlier dialog-based navigation systems [2] for the VH simulator [3], such as the one by [2], relied on semi-automatically constructed knowledge graphs and manual dialog rules, recent studies have shifted towards using LLMs for environmental perception and planning [4, 5].

A common practice in these recent systems is to embed all available environmental knowledge into the LLM's prompt. However, this approach has two major flaws: it can introduce irrelevant information that harms the accuracy of the LLM's output, and it becomes impractical and expensive with commercial LLMs that have token-based pricing models.

To address these issues, our study proposes an LLM-driven navigation system incorporating a novel prompt compression technique. In contrast to previous navigation approaches that embed all
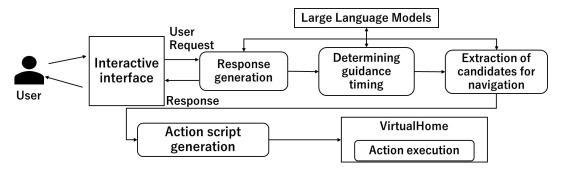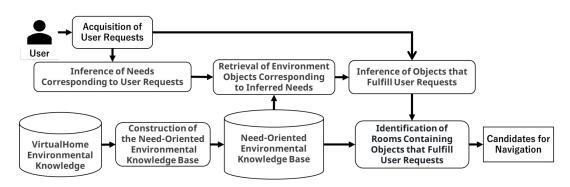
Figure 1: Diagram of the proposed system

Figure 2: Diagram of the proposed method

available environmental knowledge into prompts, our method leverages a need-oriented environmental knowledge base inspired by Murray's theory of human needs to selectively filter information. This ensures that only task-relevant context is preserved while reducing token usage and computational costs. For evaluation, our work adapts question data from the functional reasoning category of the OpenEQA benchmark [6], which involves commonsense reasoning highly relevant to navigation and decision-making tasks in household environments.

3. Method

The architecture of the proposed system is illustrated in Figure 1. It features a text-based dialog interface for user interaction. The system initiates the process by prompting a LLM to assume the role of a guide that infers user intent and facilitates a dialog. When a user issues a request involving guidance to a specific room or object, the system employs the proposed method to present a set of candidate destinations. An action script is then generated based on the selected destination, which is subsequently executed to complete the navigation.

The core of the proposed method lies within the "response generation" module of Figure 1. The internal structure is outlined in Figure 2, consisting the following steps:

1. Constructing a need-oriented environmental knowledge base for the VH environment, based on Murray's theory of human needs [7], associating each need with a relevent objects and explanatory descriptions.
2. Capturing user requests through the dialog interface.
3. Inferring the user's underlying need using a LLM.
4. Retrieving environmental objects from the knowledge base based on the inferred need.
5. Narrowing down to specific objects that satisfy the user's request through further LLM inference.
6. Identifying rooms containing the selected object(s) and presenting room-object pairs as navigation options.

This multi-stage method centers on the creation and utilization of a "Need-Oriented Environmental Knowledge Base." Inspired by psychologist Henry Murray's theory of human needs, the system selects 23 needs from the original 40, prioritizig those most relevent to household settings, such as "Thirst," "Sleep," and "Order." For each of these 23 needs, a mapping was created that associates the need with corresponding actions, a detailed textual explanation, and a curated list of VH objects that can satisfy that need. For instance, the "Thirst" need is linked to objects like "cup" and "faucet." A separate knowledge base records the object composition of each room, defining room–object relationships. To ensure computational efficiency, both knowledge bases are encoded in a custom lightweight format, rather than a standard knowledge graph representation like JSON-LD. This method minimizes the number of tokens required when incorporating this information into LLM prompts.

The full system is implemented with a Telegram bot interface. When a user sends a request, it initiates a sophisticated, multi-step inference process managed by the LLM. The first step is to infer the user's underlying need from their textual request. To accomplish this, the system was tested with two distinct prompt engineering strategies: one that prompts the LLM to use its general knowledge of Murray's theory, and in second, more verbose method that gives the LLM full textual descriptions of each need for richer semantic context. Once the need is identified, the system proceeds to the crucial prompt compression step: it retrieves only the objects associated with that specific need from the knowledge base. For example, if a user states they are hot and thirsty, only relevant objects such as "air conditioner" and "water glass" are selected, while filtering out hundreds of irrelevant items like "book" or "pillow." In the next step, a new prompt is generated, combining the original user request with the filtered object list. The LLM is tasked with identifying the single most appropriate object to fulfill the user's intent. In the final step, the system uses its room-object knowledge base to identify the location of this target object. The process concludes by presenting these identified room-object pairs to the user as candidate destinations for navigation, effectively guiding them to their goal.

4. Evaluation

4.1. Evaluation Overview

In this experiment, we evaluated the accuracy of two key processes in our system (1) Inference of needs Corresponding to User Requests and (2) Inference of Objects That Satisfy User Requests.

We used GPT-4o (gpt-4o-2025-04-17) , provided by OpenAI, as the LLMs. Accuracy was assessed using standard metrics: precision, recall, and F1-score. The evaluation was conducted using a custom dataset derived from the OpenEQA. From the original 1,636 question-answer pairs in OpenEQA, we selected 93 navigation-related questions that require commonsense reasoning. Since the OpenEQA original answers are not directly compatible with the VH environment, we conducted a questionnaire-based survey to collect VH-compatible ground-truth annotations.

4.2. Results and Discussion

The evaluation results are summarized in Table 1 and Table 2. Table 1 presents the experimental results for inferring needs corresponding to user requests. Table 2 compares the three prompting strategies in terms of navigation accuracy, average number of prompt tokens, and token count variability (standard deviation). The three strategies are defined as: (1) Murray's Theory Description prompts the LLM to infer user needs based solely on its internal understanding of Murray's psychological theory. (2) Needs and Explanations provides the LLM with explicit textual descriptions of each need. (3) All Environment Knowledge includes the complete set of environmental knowledge in the prompt without compression.

Across all settings, the best performance was achieved using the "Needs and Explanations" prompt. This approach outperformed the others in precision, recall, and F1 score, particularly when using GPT-4o, achieving a precision of 0.693, recall of 0.615, and the F1 score of 0.610. These results indicate that explicit descriptions of needs help the model better infer intent than relying on Murray's theory implicitly embedded in the model.

Table 1
Experimental results for inferring needs corresponding to user requests

Method	Precision	Recall	F value
Murray's Theory Description	0.628	0.547	0.590
Needs and Explanations	**0.693**	**0.615**	**0.610**

Table 2
Inference accuracy of items satisfying user requests, average token count, and token count standard deviation

Method	Precision	Recall	F1 Score	Token Avg.	Token Std. Dev.
All Environment Knowledge	0.540	0.493	0.467	1,091	24.154
Murray's Theory Description	0.534	0.594	0.526	**460**	56.555
Needs and Explanations	**0.590**	**0.600**	**0.548**	753	58.518

In terms of prompt compression, the proposed methods significantly reduced the number of input tokens compared to the baseline (approximately 1,100 tokens). The "Murray's Needs Theory Description" prompt required about 450 tokens and "Needs and Explanations" prompt required about 750, demonstrating a substantial reduction in input size while maintaining or improving inference accuracy. These results indicate that all proposed methods successfully reduced the average number of tokens per navigation session compared to the baseline. Notably, the method using "Murray's need theory description" yielded the greatest compression, reducing the prompt size by approximately 600 tokens relative to the full-knowledge approach. Given the evaluation dataset contains 93 instances, this equates to a total reduction of roughly 56,000 tokens across the dataset. These findings suggest that while the "Needs and Explanations" method yields the highest accuracy, the "Murray's need theory description" method offers the most efficient token usage.

5. Conclusion

Our proposed prompt compression technique, using a need-oriented knowledge base for a LLM-based navigation system, successfully improved both inference accuracy and computational efficiency. The "needs and explanations" method was the most accurate, achieving the highest F1 score of 0.548, while the "Murray's need theory description" approach was the most token-efficient. However, significant challenges remain. The underlying psychological theory is not perfectly suited for navigation, and the knowledge base requires manual construction, limiting scalability. Furthermore, the system struggles with ambiguous user requests. Future work will focus on creating a more navigation-specific need taxonomy, automating knowledge base construction, and developing robust fallback strategies.

Acknowledgments

Declaration on Generative AI

During the preparation of this work, the authors used **GPT-4o** to check grammar and improve readability. After using this tool, the authors carefully reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] J. Duan, S. Yu, H. L. Tan, H. Zhu, C. Tan, A Survey of Embodied AI: From Simulators to Research Tasks, IEEE Transactions on Emerging Topics in Computational Intelligence 6 (2022) 230–244.

[2] A. Schalkwijk, M. Yatsu, T. Morita, An Interactive Virtual Home Navigation System Based on Home Ontology and Commonsense Reasoning, Information 13 (2022).

[3] X. Puig, K. Ra, M. Boben, J. Li, T. Wang, S. Fidler, A. Torralba, VirtualHome: Simulating Household Activities Via Programs, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 8494–8502.

[4] B. Y. Lin, C. Huang, Q. Liu, W. Gu, S. Sommerer, X. Ren, On Grounded Planning for Embodied Tasks with Language Models, Proceedings of the AAAI Conference on Artificial Intelligence 37 (2023) 13192–13200.

[5] Y. Takuma, F. Jiading, L. Peng, Z. Huanyu, J. Tianchong, L. Shengjie, P. Ben, Y. David, M. Hongyuan, W. M. R., Statler: State-maintaining language models for embodied reasoning, in: 2024 IEEE International Conference on Robotics and Automation (ICRA), 2024, pp. 15083–15091.

[6] M. Arjun*, A. Ajay*, Z. Xiaohan*, P. Pranav, Y. Sriram, H. Mikael, S. Sneha, M. Paul, M. Oleksandr, A. Sergio, Y. Karmesh, L. Qiyang, N. Ben, S. Mohit, B. Vincent, Z. Shiqi, A. Pulkit, B. Yonatan, B. Dhruv, K. Mrinal, M. Franziska, P. Chris, S. Sasha, R. Aravind, OpenEQA: Embodied Question Answering in the Era of Foundation Models, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024, pp. 16488–16498.

[7] H. A. Murray, Explorations in personality, APA PsycInfo (1938).