Development of A SayCan-based Task Planning System Capable of Handling Abstract Nouns

Kosei Yamao
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu, Kitakyushu, 808-0196, Japan

Daiju Kanaoka
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu, Kitakyushu, 808-0196, Japan

Kosei Isomoto
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu, Kitakyushu, 808-0196, Japan

Akinobu Mizutani
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu, Kitakyushu, 808-0196, Japan

Yuichiro Tanaka
Research Center for Neuromorphic AI Hardware, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu, Kitakyushu 808-0196, Japan

Hakaru Tamukoh
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu, Kitakyushu, 808-0196, Japan
Email: yamao.kosei665@mail.kyutech.jp

Abstract

The task planning system is required to accomplish various requests from a human in real-world environments. SayCan, one of the task planning systems, has high accuracy. However, its accuracy decreases for requests that include abstract nouns of the ambiguous word/phrase. We propose a novel task planning system based on SayCan that introduces a function for checking concrete names of abstract nouns and a rule-based skill extraction, enhancing accuracy. The proposed system facilitates the interpretation of requests and enables appropriate task planning. The effectiveness of the proposed system was demonstrated at RoboCup@Home, where it achieved high performance.

Keywords: Task planning, Home service robot, Large language model, RoboCup@Home

1. Introduction

Recently, the demand for home service robots has been increasing due to a low birth rate and an aging population, and research on such robots has been active [1], [2], [3], [4], [5], [6]. Highly accurate task planning is necessary to realize a general-purpose service robot that performs appropriate actions in response to human requests.

SayCan, one of the task planning systems [7]. consists of two modules: a Say module and a Can module. The Say module outputs the likelihood of each skill based on the language instruction, and the Can module outputs the likelihood of each skill based on the robot's current state. It decides which skill to execute next based on the likelihood output from two modules. Although SayCan is highly accurate, the accuracy decreases for commands containing abstract nouns such as “fruit” and “drink.”

In this study, we propose a SayCan-based task planning system that introduces a function for checking concrete names of abstract nouns and a rule-based skill extraction. The contributions of this study are as follows.

- More straightforward interpretation of human requests and more accurate task planning
- Demonstrated high performance in a competition to evaluate the performance of home service robots

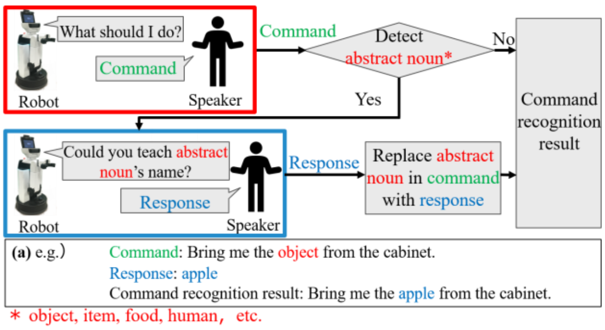

Fig. 1. Schematic of command recognition

2. Preliminaries and Related Works

2.1. *RoboCup@Home*

RoboCup@Home is an international competition for the technical development of home service robots [8]. RoboCup@Home includes two tests: General Purpose Service Robot (GPSR) and the more challenging Enhanced GPSR (EGPSR), in which robots listen to various commands and perform appropriate actions in the home environment. In competitions, commands generated from a command generator [9] are used. The command generator may output commands that include abstract nouns such as “fruit,” as in “Please find the fruit in the dishwasher.” The commands used in EGPSR are more complex than those used in GPSR and require more complex actions.

2.2. *SayCan*

SayCan has a predefined skill set that the robot can perform, and it determines which skill to execute from a given command. SayCan consists of the Say module and the Can module. The Say module uses a large language model (LLM) function to output the likelihood of each word in a sentence. The Say module inputs the given command and skill descriptions to the LLM and outputs the likelihood of each skill. The Can module acquires information about the external world from cameras and sensors and predicts which skills will most likely be executed. The results of these two modules are combined to determine the next skill to perform.

We think the Can module is difficult to implement on other robots because it is realized by reinforcement learning. SayCan is shown in the paper to be highly accurate task planning even with only a Say module without a Can module. The experimental results in the paper confirm that commands containing abstract nouns, such as “Bring me a fruit,” reduce accuracy.

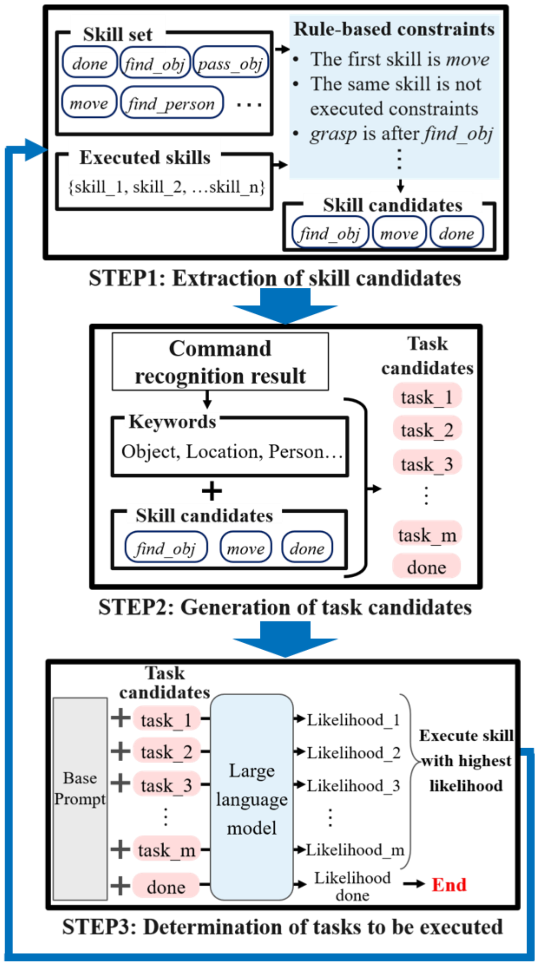

Fig. 2. Schematic of task planning

3. Proposed System

In this chapter, we describe the proposed system. The proposed system consists of command recognition and task planning modules.

3.1. *Command recognition*

Fig. 1 shows a schematic diagram of the function of command recognition in the proposed system. If an abstract noun such as an object, item, or food registered in the database is included in a given command, the proposed system asks for the concrete name of the abstract noun. This process aims to obtain a concrete name that the robot can recognize since it may not recognize the object of the abstract noun. As an example, as shown in Fig. 1(a), the robot asks for the concrete name of the abstract noun “object” and replaces “object” with “apple” depending on the result of the response.

Table 1. Skill set of the proposed system

Skill	Task
move	go to the {PLACE}
follow	follow the target
find_obj	find the {OBJECT} on the {PLACE}
find_person	find {PERSON}
observe_obj	look at the {PLACE} to check objects
observe_person	look at the {PLACE} to check people
grasp_obj	grasp the {OBJECT}
put	take the {OBJECT} to the {PLACE}
pass_obj	pass the {OBJECT}
answer_question	answer a question
say	say {LLM's output}
done	done

3.2. *Task Planning*

Fig. 2 shows a schematic diagram of the task planning function in the proposed system. The task planning consists of three steps.

STEP1: Extraction of skill candidates

Extract skill candidates from the skill set using rule-based constraints. The constraints are based on the order of skills, such as "The first skill is the move," "The same skill is not executed consecutively," and so on. This process reduces the number of LLM inputs. Eliminating skills that are very unlikely to be used is also expected to improve the accuracy of task planning. Table 1 shows the skill set of the proposed system. This skill set includes the basic skills the home service robot needs to perform various tasks. The say skill's speech content is complemented by LLM's output.

STEP2: Generation of task candidates

First, this process extracts object, place, and person names registered in the database in advance as keywords from the command recognition results. Then, generate the task candidates by combining the keywords and skill candidates. Table 1 shows the format of task candidates.

STEP3: Determination of tasks to be executed

A combination of base prompts and candidate tasks are input to the LLM, which outputs the likelihood of each candidate task. The robot then performs the task with the highest likelihood. Fig. 3 shows the base prompt. The base prompt gives the LLM information, such as the date, time, and role the command should follow.

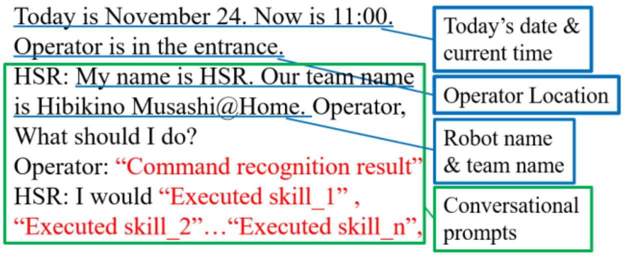

Fig. 3. Base prompt for the proposed system

4. Experiments

We conducted the following experiments to evaluate the proposed system. We use OpenAI's text-davinci-003 [10] as the LLM in the proposed system as a condition in experiments.

Experiment 1. Evaluation experiment with command output using the command generator [9]

Experiment 2. Evaluation experiment using actual competitions at RoboCup@Home 2023

In Experiment 1, we randomly generated 100 GPSR and 100 EGPSR commands, each using the command generator. In this experiment, the object and location names of the generated commands are those used in RoboCup@Home 2023. We input the generated commands into the proposed system and evaluate whether the task plan obtained as the output of the generated commands is feasible to realize the given commands. Specifically, we determined whether the sequence and targets of the skills included in the task plans can accomplish the corresponding commands.

In Experiment 2, we participated in RoboCup@Home 2023 held in Bordeaux, France, in July 2023 as Hibikino-Musashi@Home (HMA) team and evaluated the proposed system in the GPSR and EGPSR.

5. Experimental Results

5.1. *Experiment 1: Use the command generator*

As a result, we confirmed that task plans for 61 commands in GPSR and 27 commands in EGPSR were successful. In particular, we confirmed that the proposed system can successfully perform task planning by checking the concrete names of the abstract nouns in the commands to obtain the concrete names of the target objects. For example, in the command "Please bring me the fruit on the desk," the robot was checking the concrete name of the abstract noun "fruit," and the response "apple" was obtained. Table 2 shows examples of commands for which task planning was inaccurate and the possible causes. We found in many cases that task planning failures were due to incomplete skill sets and commands that did not include the required actions.

Table 2. Examples of commands that failed in task planning and possible causes

Command	Causality
Go to the desk, look for the mug, and place it on the entrance door	The command does not specify *grasp*.
Could you please close the entrance door	No skill to open/close doors.
Hand me some coke in a mug	No skill to pour liquids.

Table 3. Results of RoboCup@Home 2023

	GPSR		EGPSR	
	Team	Score	Team	Score
1st	Tidyboy	400	**HMA**	700
2nd	TRAIL	300	TRAIL	400
3rd	**HMA**	200	Tidyboy	300

5.2. *Experiment 2: RoboCup@Home 2023*

In GPSR, the command given to the robot was, "Bring me the object behind the lemon from the cabinet." In response to this command, the robot asked for the concrete name of the abstract noun "object." The operator responded, "Tropical juice," however, the speech recognition failed, and the system did not replace the abstract noun. It scored partial points because it brought the object to the operator.

In EGPSR, in response to the command, "Get acquainted with Morgan at the exit, then find him in the living room, please," the robot found Morgan at the exit. After that, the robot said to Morgan, "Hello, Morgan. I'm HSR from Hibikino Musashi@Home". Finally, the robot moved to the living room and found Morgan. This command was judged to be successful and received a score.

Table 3 shows the top three teams and their scores in GPSR and EGPSR at RoboCup@Home 2023. As a result, HMA was ranked 3rd in GPSR and 1st in EGPSR, showing that the robot works well under real-world conditions.

6. Discussion and Conclusion

In this study, we proposed a SayCan-based task planning system, which introduces a function to check the concrete name of abstract nouns and a rule-based skill extraction. We demonstrated the system's effectiveness through experiments using the proposed system and evaluations in competitions. However, the proposed system has some issues, such as the incompleteness of the skill set and the fact that task planning sometimes fails for commands that do not contain the required actions. To realize a more accurate task planning system, we think it is necessary to develop skills such as those for opening/closing doors and to devise prompts to optimize the effectiveness of LLM [11].

In addition, for the system to operate in a real-world environment, it is necessary to acquire environment-specific knowledge, such as the preferences and habits of each family member. For environment-specific knowledge acquisition and utilization, studies of episodic memory acquisition models inspired by the hippocampus are undergoing [12], [13], [14]. In the future, we will integrate the episodic memory acquisition model into the robot to operate based on previously acquired environment-specific knowledge instead of checking concrete names of abstract nouns each time.

Acknowledgments

This paper is based on results obtained from a project, JPNP16007, commissioned by the New Energy and Industrial Technology Development Organization (NEDO). This work was supported in part by JSPS KAKENHI (23H03468, 23K18495).

References

1. "Fuji Keizai Group," https://www.fuji-keizai.co.jp/report/detail.html?code=162208813, (Accessed 12/12/2023).
2. T. Ono, D. Kanaoka, T. Shiba, S. Tokuno, Y. Yano, A. Mizutani, I. Matsumoto, H. Amano and H. Tamukoh, Solution of World Robot Challenge 2020 Partner Robot Challenge (Real Space), Advanced Robotics, 2022, pp. 870-889.
3. Y. Yano, K. Isomoto, T. Ono, and H. Tamukoh, Autonomous waiter robot system for recognizing customers, taking orders, and serving food, Proceedings of the 26th RoboCup International Symposium, 2023.
4. K. Isomoto, Y. Yano, Y. Tanaka, and H. Tamukoh, Robust trash can lid opening system, Proceedings of the 2023 International Workshop on Smart Info-Media Systems in Asia (SISA), 2023.
5. T. Shiba, T. Ono and H. Tamukoh, Object Search and Empty Space Detection System for Home Service Robot, Proceedings of the 2023 International Conference on Artificial Life and Robotics (ICAROB2023), 2023.
6. D. Kanaoka, Y. Tanaka and H. Tamukoh, Applying Center Loss to Multidimensional Feature Space in Deep Neural Networks for Open-set Recognition, Proceedings of 17th International Conference on Computer Vision Theory and Applications (VISAPP2022), Vol. 5, 2022, pp. 359-365.
7. M. Ahn, A. Brohan, N. Brown, Y. Chebotar, O. Cortes, B. David, C. Finn, C. Fu, K. Gopalakrishnan, K. Hausman, A. Herzog, D. Ho, J. Hsu, J. Ibarz, B. Ichter, A. Irpan, E. Jang, R. J. Ruano, K. Jeffrey, S. Jesmonth, N. J. Joshi, R. Julian, D. Kalashnikov, Y. Kuang, K.H. Lee, S. Levine, Y. Lu, L. Luu, C. Parada, P. Pastor, J. Quiambao, K. Rao, J. Rettinghouse, D. Reyes, P. Sermanet, N. Sievers, C. Tan, A. Toshev, V. Vanhoucke, F. Xia, T" Xia', P. Xu, S. Xu,

M. Yan, and A. Zeng, Do As I Can, Not As I Say: Grounding Language in Robotic Affordances, arXiv preprint arXiv:2204.01691, 2022.

8. "RoboCup@Home." https://www.robocup.org/domains/3. (Accessed 12/12/2023).
9. "RoboCup@Home Command Generator." https://github.com/kyordhel/GPSRCmdGen. (Accessed 12/12/2023).
10. "OpenAI GPT-3.5 API [text-davinci-003]." https://platform.openai.com/docs/models/gpt-3-5. (Accessed 12/12/2023).
11. J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. V. Le, and D. Zhou, Chain-of-thought prompting elicits reasoning in large language models, Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), vol. 35, 2022, pp. 24824–24837.
12. Y. Tanaka, H. Tamukoh, K. Tateno, Y. Katori, and T. Morie, A brain-inspired artificial intelligence model of hippocampus, amygdala, and prefrontal cortex on home service robots, Proceedings of the 2020 International Symposium on Nonlinear Theory and Its Applications (NOLTA), 2020, pp. 138–141.
13. A. Mizutani, Y. Tanaka, H. Tamukoh, Y. Katori, K. Tateno and T. Morie, Brain-inspired neural network navigation system with hippocampus, prefrontal cortex, and amygdala functions, Proceedings of 2021 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), 2021.
14. A. Mizutani, Y. Tanaka, H. Tamukoh, K. Tateno, O. Nomura and T. Morie, A knowledge acquisition system with a large language model and a hippocampus model for home service robots, Proceeding of the Institute of Electronics, Information and Communication Engineers (IEICE) Rep. vol. 123, no. 208, 2023, pp. 13–18.

Authors Introduction

Mr. Kosei Yamao

He received the B.Eng. degree from the National Institute of Technology, Ube College, Japan, in 2023. He is a Master's degree student at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include task planning and home service robots.

Mr. Daiju Kanaoka

He received the B.Eng. and the M.Eng. degree from Kyushu Institute of Technology, Japan, in 2020 and 2022, respectively. He is currently a Ph.D. student at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include object recognition, multimodal learning, and home service robots. He is a student member of IEEE.

Mr. Kosei Isomoto

He received the B.Eng. degree from Kyushu Institute of Technology, Japan, in 2023, respectively. He is currently a Master's degree student at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, Japan. His research interests include brain-inspired artificial intelligence and home service robots.

Mr. Akinobu Mizutani

He received the B.Eng. and the M.Eng. degree from the Kyushu Institute of Technology, Japan, in 2021 and 2023, respectively. He is currently a Ph.D. student at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, Japan. His research interests include brain-inspired artificial intelligence and its robot application.

Dr. Yuichiro Tanaka

He received the B.Eng. M.Eng., and Ph.D. degrees from the Kyushu Institute of Technology in 2016, 2018, and 2021, respectively. He has also been a research fellow at the Japan Society for the Promotion of Science (JSPS) from 2019 to 2021. He has been an assistant professor at the Research Center for Neuromorphic AI Hardware of the Kyushu Institute of Technology since 2021. His research interests include neural networks, digital hardware implementation, and home service robots. He is a member of IEEE, IEICE, and JNNS.

Prof. Hakaru Tamukoh

He received the B.Eng. degree from Miyazaki University, Japan, in 2001. He received the M.Eng and the Ph.D. degree from Kyushu Institute of Technology, Japan, in 2003 and 2006, respectively. He was a postdoctoral research fellow of 21st century center of excellent program at Kyushu Institute of Technology, from April 2006 to September 2007. He was an assistant professor of Tokyo University of Agriculture and Technology, from October 2007 to January 2013. He had been an associate professor from February 2013 to March 2021 and is currrectly a professor in the graduate school of Life Science and System Engineering, Kyushu Institute of Technology, Japan. His research interest includes hardware/software complex system, digital hardware design, neural networks, soft-computing and home service robots. He is a member of IEICE, SOFT, JNNS, IEEE, JSAI and RSJ.