Grounding Human-Robot Dialogue in Environment Sensors to help Elderly in Activities of Daily Living

Ifrah Idrees
Dept. of Computer Science
Brown University
Providence, USA
ifrah_idrees@brown.edu

Stefanie Tellex
Dept. of Computer Science
Brown University
Providence, USA
stefie10@cs.brown.edu

Momotaz Begum
Dept. of Computer Science
University of New Hampshire
mbegum@cs.unh.edu

Abstract

The older population is set to more than double by 2050 worldwide, and conversational robots can play an essential role in improving their quality of life. In the last decade, there has been huge progress in the development of virtual healthcare assistants and social robots, with emphasis on grounding natural language commands from the human to the agent and having an engaging conversation for setting medical reminders and functioning as a companion. Conversational robot coaching dialogue systems can be used to help the elderly with Activities of Daily Living (ADLs). In this paper, we provide a proof of concept implementation of a context-aware multi-modal dialogue manager for robots that aim to help the elderly complete the ADLs safely. Our implementation interacts with the human and develops context-awareness by grounding dialogue in the sensory information received. It then fuses the current context with the user intent to learn a policy for carrying out a conversation with the elderly and guiding them in safely completing the sequential ADL task. We provide initial simulation results of our dialogue manager and show that a real physical robot can ground dialogue in information about the environment, assisting the human in completing their task of cooking safely with a success rate of 73 %.

1 Introduction

The older population, especially with dementia or cognitive impairment, has difficulties performing daily tasks, enhancing their health and psychological well-being, and physical assistance in walking or even communication [11, 17, 12]. Robots can help the elderly live in their own homes and community more independently and safely by providing physical assistance or servicing as assistive social robots [12]. This paper will focus on conversational social robots. ADLs are sequential tasks, and the sequence may vary depending on context

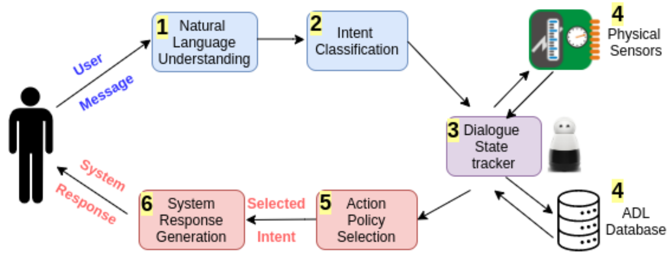


Figure 1: System architecture of our environment context-aware dialogue manager for Kuri robot

(i.e., the environment, person, the task). Therefore, a good dialogue manager for ADLs needs contextual awareness by collecting environmental information through sensors. Previously proposed dialogue managers in healthcare and social robotics lack insights about selecting the action policy for the system's response based on a combination of environmental context and dialogue context [14]. In the case of social robotics, the works focus on developing social behaviors of robots in assisting the elderly, as seen from the lab of Fasola and Matarić [6] and Kidd and Breazeal [8], with less emphasis on the language understanding part and more emphasis on the non-verbal queues.

Consequently, we advocate for and propose a context-aware, task-oriented dialogue manager for robots grounded in the elderly's health, physical condition, the surrounding environment, and the task. In particular, we will be deploying our dialogue-manager on the social and interactive Kuri robot from Mayfield [15]. We envision our dialogue manager to build context-awareness by:

- implementing a hierarchical finite state machine that supports various instructions of ADLs such as washing hands, cooking and fetching objects
- processing the sensor value of a combination of sensors to estimate the state of current step of the task on two levels - degree of safeness and completeness
- combining the dialogue history, intent of the

Robot -> Great, Next step in cooking is:
Heat the pot on high heat and bring the water to a full rolling boil
Do you want me to track to see if the activity is being completed safely or will you tell me when you are done with this step?
User input -> I will tell you when I am done
Robot -> Yes you can surely complete the activity on your own, but for your safety let me monitor you and see when the activity gets completed
User input -> hmm, okay you can monitor me
Robot -> Sure, for your safety I will monitor you and see when the activity gets completed
User input -> okay. I am proceeding with the step
Robot -> Good job, you have completed the step safely.

Figure 2: Segment of Sample Dialogue - the pink represents the user, and blue represents the robot

current dialog, with generated context to select an action for the system response.

- learning the action policy for the ADL domain dialogues by collecting a real-world corpus from simulations trials with the elderly.

In this paper, we describe our initial study of the proof of concept implementation for a context-aware task-oriented dialogue manager for robots that help the user safely complete the task of cooking. We present a qualitative evaluation of our simulation showing that with grounding dialogue in the physical sensors, it gets easier for the robot to ensure the safety of the elderly while completing the task rather than relying on self-reporting. We provide qualitative analysis with a robot demonstration and initial quantitative analysis of 26 trials of our robot dialogue system with 73% success rate.

2 Related Work

In this section, we will discuss the following three lines of work:

Task-oriented dialogues in healthcare: Research in this field has been targeted towards helping patients with clinical decision support, reducing stress, or setting medicine reminders [5, 10]. These works focus on building systems in typical social roles to realize natural conversations, incorporating self-reported health monitoring outcomes to update the models. These work lack insights in combining health monitoring aspects of the elderly through physical sensors with a spoken dialogue regarding completing the activities of daily living.

Social Robotics: A great deal of research has been done on developing social behaviors for the robots to assist autistic children in cognitive training, helping users reduce weight and psychological therapy and cognitive training of the elderly. These works mostly focus on user motivation and personalization to improve user experience [13, 8, 1]. However, in most of the studies, the social robot is hard-coded with conversational cues. Their research focuses on how to make communication better with other non-verbal cues such as embodiment, gesture, and measure its effectiveness on the users such as the robot exercise coach [6]. The dialogue managers for social robots take into account the physical context of the environment. However, their focus is not on combining this context with a dialogue manager to learn an action policy to generate a natural language response. There has been separate work in object search to help elderly in case they forget the location of objects but with no emphasis on human-robot interaction [7] . [16] describes a multi-year development of prototype to support agents with ongoing engagement and dialog for companionship purposes. Insights from this study can be incorporated for our complete version to help elderly with ADL.

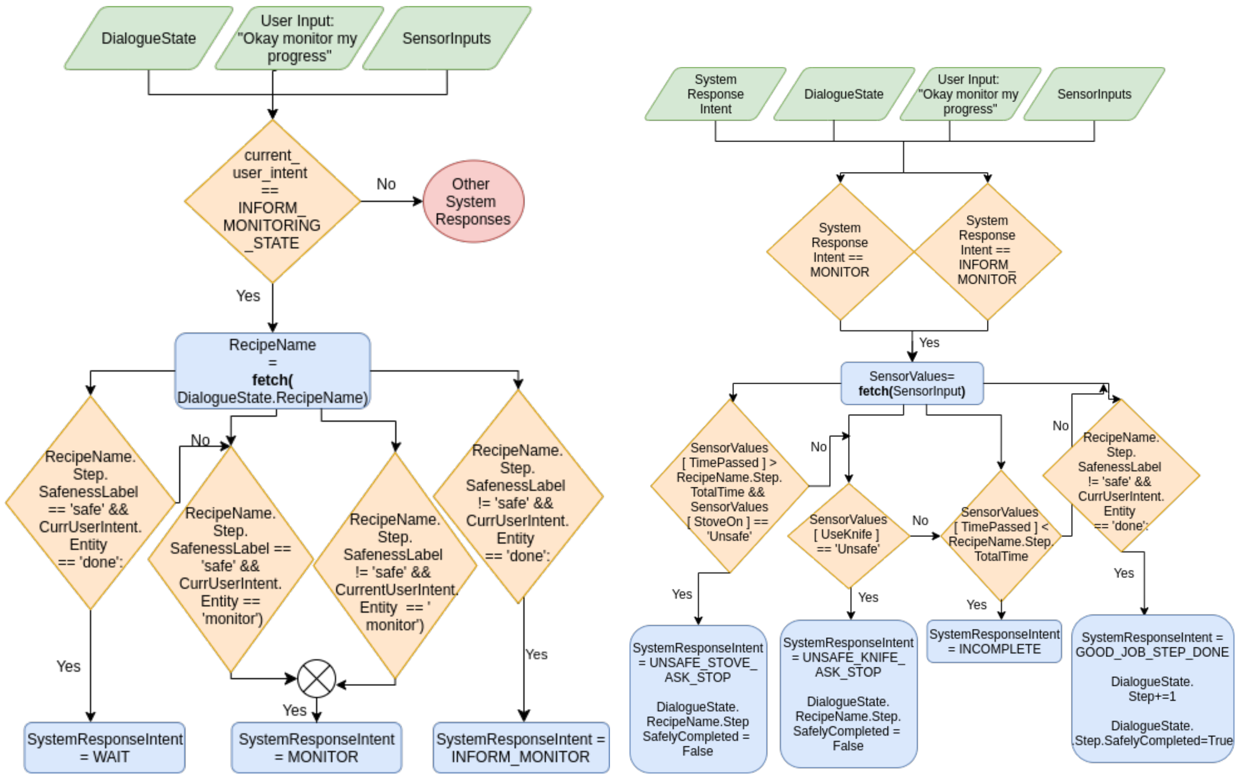
Conversational Social Robots: Relevant examples include Mabu by Cory Kidd [4] and Koichiro et al. [9]. The former dialogue manager checks in on the health of the elderly, addressing challenges to treatment adherence. However, the input from the user is not a natural language utterance; rather, users select input from a set of provided prompts. On the other hand, the latter builds a dialogue manager to handle the loneliness of the elderly and is not task-oriented or targeted to the ADL.

3 System Architecture

Fig-1 shows the concept diagram of our context-aware, task-oriented dialogue manger for robots. We build the architecture of our initial prototype, over the base chatbot architecture of open source RASA [2]. In our prototype, the **Modules 1, 2 and 3** of Fig-1 are off-the-shelf components as provided by RASA. We use white space tokenizer combined with supervised word embedding for extracting lexical, syntactic features, and entities from the user messages for natural language understanding. While DIETClassifier (Dual Intent and Entity Transformer) is used for Intent classification and slot-filling architecture for state representation and tracking [3].

Our main efforts were two-pronged:

1. **Module 3** - implementing a finite state machine over the slot-filling architecture to allow our dialogue manager to update the dialogue state as the task proceeds. For this, we used customs actions API of RASA. In our prototype, the only state machine is for the cooking activity.



(a) (b)

Figure 3: Logic behind creating context awareness dialogue manager

2. **Module 4** - making our dialogue manager context-aware to update its dialogue state based on user input and also physical context - task, physical environment, elderly's condition. The logic behind how we combine the dialogue state with physical sensor values is presented in Fig-3. As the user inputs a message, the intent of the current user input and the dialogue state decide whether the physical context needs to be invoked. Our algorithm then exhaustively conditions over the possible states of the sensor values to decide the system's response. In our case, cooking recipe steps are fetched from stored records. Each step has been pre-labeled as inherently safe or unsafe. This information is combined with the $CurrUserIntent$ of whether they want to be monitored for help and with $SensorValues$ to determine whether the state of the recipe step is a) complete, b) incomplete (for time-constrained steps), and c) safe or d)unsafe (for all the steps). Since we are not using actual sensors in the prototype, the sensor values for states $StoveOn$, $UsesKnife$, $TimePassed$ are stored as binary values in a $data\_file$ containing all possible feasible permutations. For a given recipe step for which the sensor values need to be processed, these values are fetched randomly from $data\_file$.

Module 5 in our prototype has an action policy is learned using a combination of i) Transformer Embedding Dialogue (TED) - a self-attention mechanism over the sequence of dialogue turns, ii) mapping policy where 21.4% of user intents are mapped to actions and iii) memoization policy where the policy predicts action if the exact conversation is in training data. We trained our dialogue manager with synthetic data for now and intent to run simulation trials with the elderly, to build a more accurate corpus. For **Module 6**, RASA takes as input the action intent from **Module 5** and generates a system's response from set of responses.

4 Evaluation

We perform both qualitative and quantitative evaluation of our system and also provide a robot demonstration. The user in our scenario intends to complete their task of cooking (one of the ADL activities), and the robot acts like a cooking coach and a health pal who gives instruction to the user to complete the task while ensuring their safety.

Sample Dialogue: Fig-2 shows an excerpt from the sample conversation of the user with our dialogue manager. The excerpt illustrates some of the useful features of our system - support for ADL domain and task-oriented, context awareness dia-

Sensor Values			System Response	Trials	Successes	Errors
Stove	Knife	Time				
0	0	$< T^*$	Some Time is left for step to complete.	4	2	2
0	0	$> T^*$	Good job, you have completed the step safely.	4	4	13
1	0	$< T^*$	action_listen	3	3	0
1	0	$> T^*$	Good job, you have completed the step safely. However, you forgot to turn the stove off.	7	6	8
0	1	NA	Knife seems to being mishandled. Please stop to avoid any injury.	4	2	2
1	1	NA	Situation seems risky. Please stop your activities.	4	2	2

Table 1: Quantitative evaluation of our prototype.
[1] T* is total time required for time constrained recipe steps.

logue for sequential tasks. During a complete conversation, our dialogue manager is shown to handle specific greeting behavior, provide instructions for the sequential task, and monitor their progress and safety if requested by the user or deemed necessary by the robot. This can be seen in the robot demonstration video (linked below) as well. If our dialogue manager did not have the context-awareness feature, the robot would have to rely upon self-reporting from the user solely. This would not have proven effective for the elderly with dementia or limited cognitive abilities since dementia affects the brain's ability to sequence, plan, and organize multiple-step activities.

Robot Demonstration: A video of a demonstration of our dialogue manager deployed on Kuri robot can be seen at https://youtu.be/fnh5vnlP4t8.

Quantitative Evaluation: We perform an initial quantitative study of our robot deployed prototype over the trial of 26 complete dialogues.

Table-1 shows a distribution of the sensor input conditions. In the Table-1 Sensor input a) "Stove" symbolizes the boolean value 1 of the stove being left on after the step is completed, b) "Knife" represents the boolean value "1" for a knife being used unsafely and c) "Time" signifies the time spent on the current recipe step. The table then shows the response of our dialogue manager in each of the sensor input conditions. We performed an analysis of the reasons behind a) unsuccessful dialogue conversations (the task of cooking not completed) and b) situations where the dialogue manager did not act as anticipated causing minor inconveniences but was still able to complete the task. The errors for both these scenarios can be categorized into a) Speech Recognition (33%) b) Intent Classification(44%) c) Action policy selection(23%). These errors did not always lead to unsuccessful conversations, as one can see in the second row of Table-1.

5 Discussion

For our prototype, we have considered only binary sensor values - 'safe' and 'unsafe' for stove and knife usage. Future integration of a multitude of sensors will be more complex. The sensors we envision to incorporate are not just limited to the robot's sensors, such as RGB-D cameras for activity and object recognition. We also hope to incorporate smart home devices such as temperature sensors for the stove, which will give us a better representation of the environment's state, allowing usage of simpler threshold-based sensor processing techniques. However, the sensors' processing needs to be time-aware with the dialogue state tracking mechanism and hence real-time. Our current implementation of a prototype dialogue manager has a finite state machine for just cooking . In the complete version, we hope to support various types of ADL, such as washing hands, cooking, and fetching objects with a hierarchical finite state machine. We hope to experiment with different state machines' structures to ensure that the performance of state tracking does not adversely affect the system's response time with an increasing number of tasks and objects.

6 Conclusion

With the design choices described in this paper, we show the feasibility of a task-oriented ADL based context-aware dialogue manager for robots. With our dialogue manager, the robot was able to assist the user in completing the cooking recipe successfully and safely with a 73% success rate. There is still significant room for improvement for our full system. We intend to achieve real-time sensor fusion for context-awareness, and to explore other data-driven approaches as well - transformer, LSTM, and Deep-RL to improve intent classification and action policy selection.

References

[1] Jordan Abdi, Ahmed Al-Hindawi, Tiffany Ng, and Marcela P Vizcaychipi. 2018. Scoping review on the use of socially assistive robot technology in elderly care. *BMJ open*, 8(2).

[2] Tom Bocklisch, Joey Faulkner, Nick Pawlowski, and Alan Nichol. 2017. Rasa: Open source language understanding and dialogue management. *arXiv preprint arXiv:1712.05181*.

[3] Tanja Bunk, Daksh Varshneya, Vladimir Vlasov, and Alan Nichol. 2020. Diet: Lightweight language understanding for dialogue systems. *arXiv preprint arXiv:2004.09936*.

[4] Gary Arnold Brian Mirletz Brien Voorhees Cory Kidd, Devon Edwards. Method and system for patient engagement.

[5] David DeVault, Ron Artstein, Grace Benn, Teresa Dey, Ed Fast, Alesia Gainer, Kallirroi Georgila, Jon Gratch, Arno Hartholt, Margaux Lhommet, et al. 2014. Simsensei kiosk: A virtual human interviewer for healthcare decision support. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pages 1061–1068.

[6] Juan Fasola and Maja J Matarić. 2013. A socially assistive robot exercise coach for the elderly. *Journal of Human-Robot Interaction*, 2(2):3–32.

[7] Ifrah Idrees, Steven P Reiss, and Stefanie Tellex. 2020. Robomem: Giving long term memory to robots. *arXiv preprint arXiv:2003.10553*.

[8] Cory D Kidd and Cynthia Breazeal. 2008. Robots at home: Understanding long-term human-robot interaction. In *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3230–3235. IEEE.

[9] Yoshino Koichiro, Murase Yukitoshi, Lubis Nurul, Sugiyama Kyoshiro, Tanaka Hiroki, Sakti Sakriani, Takamichi Shinnosuke, and Nakamura Satoshi. 2019. Spoken dialogue robot for watching daily life of elderly people. In *International Workshop on Spoken Dialogue System Technology*.

[10] Andreas Komninos and Sofia Stamou. 2006. Healthpal: an intelligent personal medical assistant for supporting the self-monitoring of healthcare in the ageing society. In *Proceedings of UbiHealth*. Citeseer.

[11] Mikaela Law, Craig Sutherland, Ho Seok Ahn, Bruce A MacDonald, Kathy Peri, Deborah L Johanson, Dina-Sara Vajsakovic, Ngaire Kerse, and Elizabeth Broadbent. 2019. Developing assistive robots for people with mild cognitive impairment and mild dementia: a qualitative study with older adults and experts in aged care. *BMJ open*, 9(9):e031937.

[12] Ester Martínez and Angel P. del Pobil. 2018. Personal robot assistants for elderly care: An overview. In *Personal Assistants: Emerging Computational Technologies*, pages 77–91. Springer.

[13] Maja Matarić. Maja j matarić - publications.

[14] Muhammad Asif Razzaq, Wajahat Ali Khan, and Sungyoung Lee. 2017. Intent-context fusioning in healthcare dialogue-based systems using jdl model. In *International Conference on Smart Homes and Health Telematics*, pages 61–72. Springer.

[15] Mayfield Robotics. 2018. Meet kuri! the adorable home robot.

[16] Candace L. Sidner, Timothy Bickmore, Bahador Nooraie, Charles Rich, Lazlo Ring, Mahni Shayganfar, and Laura Vardoulakis. 2018. Creating new technologies for companionable agents to support isolated older adults. *ACM Trans. Interact. Intell. Syst.*, 8(3).

[17] Healthcare Sova. 2018. How does dementia affect everyday life?