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 (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
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 dialogue 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 manager 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.

<table>
<thead>
<tr>
<th>Robot</th>
<th>Great, Next step in cooking is:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Heat the pot on high heat and bring the water to a full boil</td>
</tr>
<tr>
<td></td>
<td>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?</td>
</tr>
<tr>
<td>User input</td>
<td>I will tell you when I am done</td>
</tr>
<tr>
<td>Robot</td>
<td>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</td>
</tr>
<tr>
<td>User input</td>
<td>I am, okay you can monitor me</td>
</tr>
<tr>
<td>Robot</td>
<td>Sure, for your safety I will monitor you and see when the activity gets completed</td>
</tr>
<tr>
<td>User input</td>
<td>okay, I am proceeding with the step</td>
</tr>
<tr>
<td>Robot</td>
<td>Good job, you have completed the step safely.</td>
</tr>
</tbody>
</table>
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 intend 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-
**Table 1: Quantitative evaluation of our prototype.**

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

1 T* is total time required for time constrained recipe steps.

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


