Dialog Management for a Social Assistive Robot in the Domain of Elderly Care

Berardina De Carolis, Giampaolo Flace, La Forgia Angela, Nicola Macchiarulo and Giovanni Melone

Department of Computer Science, University of Bari Italy
Exprivia S.p.A, Molfetta, Italy

Abstract

The world population is aging and there are main concerns of the aged care industry to provide appropriate care for elderly people as their health and independent functioning decline. Socially Assistive Robot technology could assume an important role in health and social care to meet this higher demand. In this paper, we present an architecture to support believable conversation capabilities of a social robot in the context of daily task assistance to elderly people. The conversational system is based on a BDI architecture that allows to mix deliberative and reactive reasoning in order to determine, at each step, which goals are valid and, consequently, which action to perform to reach a goal according to the current dialog context, including the emotional state of the user. The architecture has been preliminary tested on a low cost robot, *Ohmni* in this case, usually dedicated to telepresence and the results of this evaluation show that, from the dialog management point of view, it is robust enough to handle dialogs typical of the elderly care domain.

Keywords

Social Robot, Conversational System, Belief Desire Intention

1. Introduction

In today's society, the geriatric population is steadily increasing and requires more and more assistance and monitoring [1]. These needs are sometimes not fully met due to the shortage of elderly care personnel. In this context, one of the most ambitious challenges is to introduce *Social Assistive Robots* (SAR) to provide assistance to elderly relying on social interaction [2].

Then, the ability to communicate using natural language and high level dialog management is a fundamental requirement for a social robot since spoken dialogue is generally considered as the most natural way for human-robot interaction [3].

Research in the field of Human-Robot Interaction (HRI) is increasingly focussing on the development of robots equipped with intelligent conversational abilities by developing dialogue systems able to deal with real-world scenarios supporting specific tasks especially in social contexts.

Simulating human-to-human communication to enhance and ease human-to-machine communication is still a challenging task, especially when the goal is to enable a natural, adaptive, context and affect aware interaction.

To this aim, we developed an architecture of a Conversational System (CS) that is able to support believable conversations in the context of daily task assistance to elderly people.

In this work, a general architecture to implement a CS is presented, with the aim of supporting the conversational capabilities of social robots, even a low cost robot like the one used for developing the prototype presented in this paper: *Ohmni*, a low cost domestic robot with Android operative system, developed by *OhmniLabs* and usually dedicated to telepresence. It does not have a humanoid appearance, but is composed of a simple metal rod that supports a microphone, a speaker, a webcam and a touchscreen. It also comes with a mobile base that allows its movement. Such very rudimentary features, and the lack of conversational capabilities or SDK functions that permit its implementation, make Ohmni a perfect candidate for using the developed conversational system.

The proposed architecture is based on the Beliefs, Desires, Intentions (BDI) one. As explained in [4], BDI architectures are primarily focused on practical reasoning, i.e. the process of mixing deliberative and reactive reasoning in order to determining, at each step, which goals are valid and, consequently, which action to perform to reach a goal according to the current state of the world. Mixing reactive/proactive approaches enables the management of coherent conversations while still being responsive to unexpected user inputs.

The dialog module, then takes into account the agent's knowledge (the beliefs) and selects at each step the best goal to be fulfilled through the activation of an intention that the robot commits to. Intentions are then satisfied through the execution of plans that may include dialog acts and service activation. The model integrates a deliberative and a reactive components since goals are triggered, revised dynamically and satisfied by triggering appropriate plans, so that an appropriate response can be generated and returned to the user.

During the interaction between users and the robot, perceptions, as in general for conversational interfaces, are captured by analysing the speech of the user during the dialogs.

The voice captured through the robot's microphone is used both for converting the *Speech To Text (STT)* and for emotion analysis. The converted text is used by *Natural Language Processing (NLP)* techniques, in order to extract user intention (intent), significant entities, and sentiment polarity. The voice signal is then analysed with a specific software to recognize emotions. These information, also called extracted perceptions, along with background knowledge and the active goal, are interpreted as beliefs and used by a dialogue management component to trigger goals that will be satisfied by pre-defined plans, which are generally not very long as they have to adapt to the unpredictable interaction of the user, who after all, can say anything.

This architecture has been fully implemented and tested on the Ohmni robotic platform in the elderly care interaction scenario realized for the SI-Robotics project.

The remaining of the paper details this conversational architecture. The next section introduces the Italian project SI-Robotics. Section 3 provides an overview of the architecture as well as the dialog system that we have developed for our robot, going into details about its components. Section 4 provides a practical application of the conversational system in the context of elderly assistance. Sections 5, finally, summarise our main contributions and restate the key challenges that human–robot interaction brings to Artificial Intelligence.

2. SI-Robotics

SI-Robotics is an Italian project that aims to design and develop innovative collaborative robotics solutions to support human beings in health and care services, in home, residential, and hospital environments. This project aims to produce advanced models of interaction, designed to motivating an active aging. These solutions are easily adaptable and allow to help elderly people in daily activities, anticipating their needs, and offering teleassistance, monitoring, and coaching services. To meet these needs, SI-Robotics proposes the realization of a cognitive agent able to interface with humans, IoT devices, social robots, and services present in the cloud [5]. The development of the SI-Robotics project has been entrusted to Exprivia, which collaborates with sixteen other partners to design the system architecture, software services, and AI-based functionalities, also taking care of their integration.

3. An Overview of the Conversational System

The Conversational System (CS) allows the social robot, Ohmni in this case, to operate as a virtual assistant, providing it with both responsive and deliberative capabilities based on the architecture illustrated in Figure 1. It is composed by two main components:

- The *Conversational User Interface (CUI)*, developed as a web application, running inside the robot's tablet, to enable interaction between the senior and Ohmni;
- The *Conversational Agent (CA)*, that based on the input received from the CUI, interprets and processes an answer, along with a possible action to be performed.
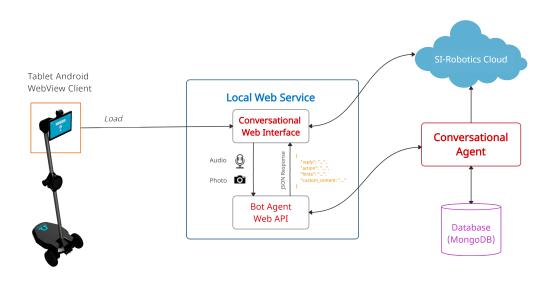

Figure 1: Conversational system architecture in the form of components and message flow.

Figure 1 shows at a high level the interaction among these components. In particular, the touchscreen on the Ohmni robot allows to manage the interaction through the CUI that has been developed as a web application due to the constraints of the Ohmni management system [6]. It is displayed in a special Android app that implements a WebView. The CUI allows not only to display multimedia content but also to manage the speech-based interaction with the user. The workflow of the interaction process will be described in details in the following section. The robot communicates with a cloud platform to recieve some events, such as the start of daily check-up or reminder task. The events are read by the CUI in order to notify the CA and a specific behavior is activated. Once the user's voice is recorded or an event is triggered, the CUI sends a request to the web API related to the CA in order to obtain a response in JSON format containing the following information: textual response to be played back to the user, one or more suggestions about what the robot expects the user to respond to, an HTML content to be rendered, and a possible action to be performed.

To provide a more general view of the components and subcomponents that make up the CS, Figure 2 shows the component diagram.
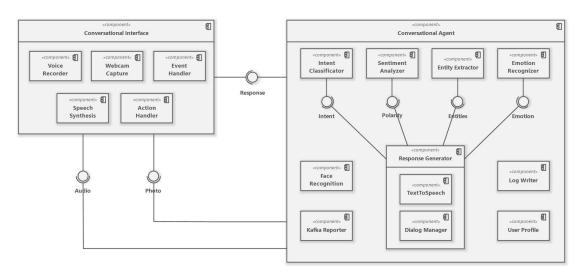

Figure 2: Conversational system architecture in the form of components and message flow.

3.1. The Conversational User Interface

For reasons strictly related to integration with the Ohmni robot, the CUI, shown in Figure 3, is implemented as a web application, using Javascript as programming language.
Its features include:

- Recording the user's voice while talking;
- Communication with the bot by submitting recorded audio or captured photo;
- Displaying the content of the CA response message (composed from: one textual answer, one or more suggestions, and possibly a multimedia content);
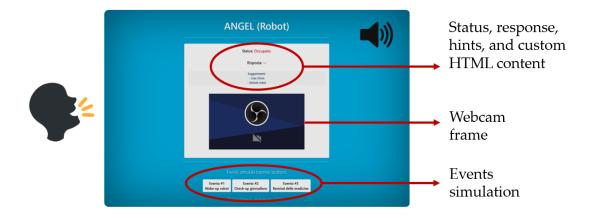
Figure 3: Conversational user interface implemented in the SI-ROBOTIC project.

- Speech synthesis to reproduce the textual response received from the bot;
- Visualization of the webcam video stream in a dedicated frame;
- Capture a photo using the robot's webcam;
- Communication with the cloud platform for receiving events;
- Execution of specific actions.

In particular, the HarkJS library is used for voice detection, while *Ohmni Standalone* library provided by Ohmnilabs is used for speech synthesis.

3.2. The Conversational Agent

The CA implements the following functionalities:

- Conversion of received audio to text *(Speech to Text)* using the *Google Speech* service;
- Facial recognition of a person and of his/her gender from the face;
- Sending communication messages to the cloud platform;
- Storing and retrieving user data from the database;
- Extracting perceptions from text and audio;
- Managing the conversational flow through a dialog management function.

The last two aspects are explored in more detail in the subsequent sections.

3.2.1. Extraction of perceptions

The following perceptions are extracted during the dialog:

- *User intention*: goal or activity the user wishes to accomplish;
- *Sentiment polarity*: positive, negative, or neutral;
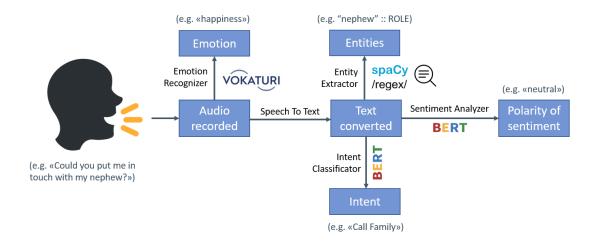- *Entities*: useful information that can be found in a sentence;

Figure 4: Extraction of perceptions with a practical example.

- *Emotion*: happiness, fear, anger, disgust, sadness, surprise.

As shown in Figure 4, the first three perceptions are extracted using the converted text, while the emotion is detected directly from the analysis of the recorded audio using Vokaturi [7]. To limit as much as possible the use of paid cognitive services on cloud, ad-hoc predictive models based on deep learning are trained for intent, entity, and sentiment recognition. In particular, intent and sentiment are predicted using *Google BERT* [8], while entities are extracted using a mixed approach based on *Spacy* [9], regular expressions, and text search considering also lemmas. Finally, the *Vokaturi* library is used for emotion recognition.

It should be noted that the use of *BERT* for intent recognition and sentiment analysis is justified by the fact that this model based on deep learning adopts a bidirectional architecture of transformers. Due to the use of the multi-head self attention mechanism, during term processing, BERT reads the entire word sequence at once, and this enables learning of contextual relationships between words (i.e. take into account the context of the sentence). In addition, it is one of the first models to adopt the technique of transfer learning within NLP. The pre-training sentence involves unsupervised learning by training the network on generic tasks, while in the fine-tuning phase, the network is reused and adapted to perform a more specific task by adopting a supervised approach. The downstream task used for both intent recognition and sentiment analysis during the fine-tuning process was sequence classification. The advantages of using transfer learning are: reduced training time, improved prediction, and the ability to use a relatively small dataset and still get good results.

3.2.2. Dialogue Management

Once perceptions are obtained, a dialogue-based mechanism, inspired by the BDI model, is used to obtain the response.

As in our approach, the conversational capabilities and the dialog management of an agent can be implmented through the BDI agent model [10] that has been used successfully in a

range of applications. This model allows for a mix of reactive behaviour and goal-directed reasoning and this supports different means for achieving a goal depending on the context and other factors [11]. The mixed reactive/proactive model enables the management of coherent conversational activity while still being responsive to unexpected user input and aware of changes in the conversation context. BDI plans provide knowledge of how to perform different types of conversational activities, while appropriate Knowledge Bases (KBs) contain information about entities associated with those activities.

BDI agent-based approaches to dialogue management have been previously proposed (e.g.,[12, 13]); however, these have typically been for task-oriented conversations (e.g.,accessing email or managing an appointment). In our approach we use the BDI framework to provide variability in the way a goal is progressively achieved, as well as in the conversational content [14].

The goals of the CA are activated on the bases of its belief, that are eventually revised according to changes in its perceptions. Goals are fulfilled through the selection of different plans whose execution is, in turn, adapted to different contexts (i.e. recognized emotions, time of the day, etc.) thus behaviors that are appropriate to the user's situation.

A goal is implemented through a parameterized function with: the converted text, the extracted perceptions, and a reference to the CA instance. The plans associated with this function are created by placing logical conditions on the values of perceptions, especially on intent. Goals are managed using a stack, so the function associated with the goal at the top of that stack is called to obtain the response.

This whole mechanism needs at least a goal function to work. For this reason, *General* goal has been implemented and placed by default as first element of the stack. In the case where there are no feasible plans (i.e. no condition in the goal function is satisfied), the CA initially responds in a generic way by notifying to the user that it has not understood and invites the user to repeat. In case this happens a second time, it will formulate the previous question again. While the dialog evolves, goals are put in the stack and the user's request is managed by the goal function.

Finally, to prevent the robot from responding when it is not needed, communication with the CA is initially disabled and then re-enabled by the user by uttering one of the following wake-words: *robot* or *ohmni*.

4. An Example in the context of Elderly Assistance

The conversational capability of the Ohmni robot have been finalized in order to support the following tasks:

- Registration of the senior profile;
- Communication of relevant symptoms and general conditions;
- Telepresence;
- Information provision (i.e. weather);
- Medicine Remind;
- Daily Check-up.

To this aim, we trained our system to recognize 10 main intents and 38 smalltalks [15]. Examples of intents that can be recognized during the conversation are the following:

- CheckUp (i.e. "let's make the daily check-up");
- RemindMed (i.e. "please remind me the pills I have to take today");
- HealthState (i.e. "I feel like I have fever");
- BotAbility (i.e. "what are you able to do?").

As far as entities are concerned, until now we are able to recognize the following 8 categories of entities:

Entity Name	Approach	Description
SYMPTOM	Spacy	Symptom or illness (e.g. fever, stomach ache)
LOC	Spacy	Country, city, or place (e.g. Mola di Bari, Molfetta)
EMOTIVE	Search	Emotional state of the user (e.g. happy, sad)
MEASUREMENT	Search	Vital parameters (e.g. pulse, blood pressure)
DAY	Search	Day of the week (e.g. Monday, Saturday)
TEMPORAL	Search	Time indicator (e.g. today, tomorrow)
ROLE	Search	Role held (e.g. doctor, nephew)
NUM	Regex	Numbers (e.g. 82, 180)

Note that while the entity LOC is recognized by the basic Italian model provided by Spacy, the entity SYMPTOM is trained manually. In addition, it was possible to use regular expressions to capture the NUM entity since the Google STT tool used automatically converts the numbers spoken by the user into numeric format.
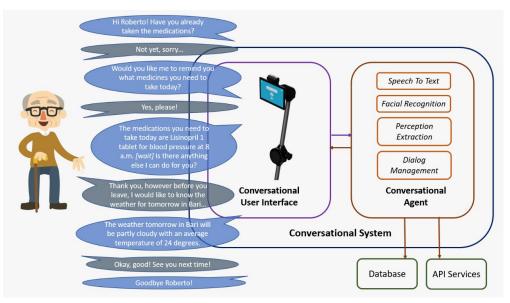

Figure 5: Example of a conversation between an elderly person and the robot. The robot receives the medicine reminder event from the cloud platform, and starts the conversation.

With the objective to clarify how the system is expected to dialog with seniors in the context of assisted daily life, the following example in Figure 5 is given. The figure describes a particular

Figure 6: Examples of CUI responses during dialogue.

conversation that could be performed by the system. Instead, in Figure 6, screenshots of the CUI are shown during the dialogue between elder and robot.

As shown in Figure 7, testing of the conversational solution was conducted both from PCs via browser and on the Ohmni robot in the Exprivia enterprise. COVID-19 restrictions prevented evaluation in ALHs (Assisted Living Houses) so that testing could be conducted with elderly people.
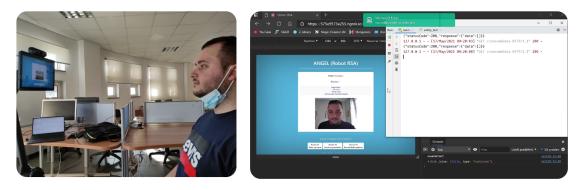

Figure 7: Testing of the conversational system from the browser and on the Ohmni robot.

5. Conclusions and Future Work

The work described in this paper is a preliminary step towards a general architecture of a Social Assistive Robot for elderly people living in Assisted Living Houses. In this phase of our research we developed a conversational system based on a BDI architecture and we integrated it into the Ohmni robot, that is a low-cost product designed only for telepresence tasks, in order to endow it with conversational capabilities. The result of the prototype version obtained so far is

satisfactory, but needs improvement especially in voice detection.

Future developments will focus on the following points:

- Add new assistive features by introducing more intents and plans;
- Adopt smart microphones capable of filtering noise to improve speech recognition;
- Evaluate the conversational system by carrying out usability and technology acceptance tests with real users (i.e. elderly people in ALHs);
- Develop services in such a way that the conversational system can also be used on other platforms.

Acknowledgments

References

[1] Elderly population across EU regions, 2020. URL: https://ec.europa.eu/eurostat/web/products-eurostat-news/-/DDN-20200402-1.

[2] D. Feil-Seifer, M. Matarić, Defining socially assistive robotics, volume 2005, 2005, pp. 465 – 468. doi:10.1109/ICORR.2005.1501143.

[3] T. Fong, I. Nourbakhsh, K. Dautenhahn, A survey of socially interactive robots, Robotics and Autonomous Systems 42 (2003) 143–166. doi:10.1016/S0921-8890(02)00372-X.

[4] A. M. Andrew, Reasoning about rational agents, by michael wooldridge, mit press, cambridge, mass., 200, xi 227pp., isbn 0-262-23213 (hardback £23.50), Robotica 19 (2001) 459–462. doi:10.1017/S0263574701213496.

[5] Si-robotics, 2021. URL: https://www.exprivia.it/it/show-info-event-full.php?id_event=6009.

[6] Webapi - ohmni developer manual, 2021. URL: https://docs.ohmnilabs.com/webapi.

[7] J. Garcia-Garcia, V. Penichet, M. Lozano, Emotion detection: a technology review, 2017, pp. 1–8. doi:10.1145/3123818.3123852.

[8] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, arXiv (2018). URL: https://arxiv.org/abs/1810.04805v2. arXiv:1810.04805.

[9] M. Honnibal, I. Montani, S. Van Landeghem, A. Boyd, spaCy: Industrial-strength Natural Language Processing in Python, 2020. URL: https://doi.org/10.5281/zenodo.1212303. doi:10.5281/zenodo.1212303.

[10] A. S. Rao, M. Georgeff, BDI Agents: From Theory to Practice, undefined (1995). URL: https://www.semanticscholar.org/paper/BDI-Agents%3A-From-Theory-to-Practice-Rao-Georgeff/8bb51f40236fd06406f22b31fcacb381539c3bf9.

[11] L. Padgham, M. Winikoff, Developing Intelligent Agent Systems: A Practical Guide, Wiley, Hoboken, NJ, USA, 2004.

[12] A. Nguyen, W. Wobcke, An agent-based approach to dialogue management in personal assistants, International Conference on Intelligent User Interfaces, Proceedings IUI (2005) 137–144. doi:10.1145/1040830.1040865.

[13] J. van Oijen, W. A. van Doesburg, F. Dignum, Goal-Based Communication Using BDI Agents as Virtual Humans in Training: An Ontology Driven Dialogue System, ResearchGate (2010) 38–52. doi:10.1007/978-3-642-18181-8_3.

[14] W. Wong, J. Thangarajah, L. Padgham, Flexible conversation management using a bdi agent approach, volume 7502, 2012, pp. 464–470. doi:10.1007/978-3-642-33197-8_48.

[15] Microsoft, Bot Frameowrk Tools, 2021. URL: https://github.com/microsoft/botframework-cli/blob/main/packages/qnamaker/docs/chit-chat-dataset.md, [Online; accessed 11. Nov. 2021].