FAIM/ISCA Workshop on Artificial Intelligence for Multimodal Human Robot Interaction
14-15 July 2018, Stockholm, Sweden

AI-based Dialogue Modelling for Social Robots

Kristiina Jokinen
AIRC, AIST Tokyo Waterfront, Japan
Kristiina.Jokinen@aist.go.jp

Abstract

The paper addresses various issues related to situated interaction between a robot agent and a human user. The focus is on dialogue modelling and implementation of a speech-based multimodal dialogue management model in a robot agent, with the aim to enable the agent to conduct natural type of dialogue interactions. The topics related to the type and representation of knowledge and the selection of appropriate knowledge as the focal point of the dialogue are discussed, considering the robot's dual characteristics as a powerful computer and a communicating agent.

1 Introduction

The need for interactive systems in various service sectors increases as digitalization proceeds in all realms of society. We are familiar with banking and administrative services, not to mention various mobile apps and social media. IoT technology can provide innovative solutions that transform services into more effective and personalized services, while health-care and elder-care support systems are under intensive study to tackle challenges caused by social and demographic changes in the society, such as increasing number of elderly people and shortage of caretaking staff [6].

In this context, it is important that intelligent devices and services are equipped with a capability to interact with humans in a manner which is efficient, flexible, informative, easy, and pleasurable, i.e. *natural* from the point of view of the user. Such interactions are best conducted via natural language dialogues which enable users to get their task completed quickly and flexibly, without needing to waste time wondering how to operate the system or what kind of commands to use.

Dialogue management technology (see an overview in [4]) has matured to the level of speech-based interactive systems being functional and useful. Question-answering systems can provide accurate and helpful answers to user' questions and many are commercially available; for instance, the most famous ones, IBM Watson, Google Home, Apple Siri and Amazon Alexa enable intelligent question-answering and chat-like conversations with users. Moreover, when spoken dialogue systems are combined with humanoid robots, which can move and insert their presence in the 3-dimensional environment, such natural language interactions can offer an intuitive interface which supports agent-like communication instead of just manipulation of a tool.

The view of an automated system as an intelligent agent can be related to *affordance*, the concept brought to HCI by [10] and suggested for natural language interactive systems by [3] and further applied in robot design e.g. by [8][9]. It concerns the system's properties or functionalities that readily suggest an appropriate way to interact with the artefact. For instance, communicative competence of a system affords natural language interaction and lends itself to intuitive use of the system. An interface with natural language communication capability enables users to utilize their knowledge of human interactions and multimodal signals, and mimic social communication in their interactions with intelligent agents.

In order to develop such interfaces, rich knowledge of the context and environment in which the dialogues take place is required. There is intensive research and development of end-to-end dialogue-systems based on big data and neural learning techniques, and questions concerning interpretability, deep (latent) semantics, and integration of common sense knowledge are some of the important issues to be addressed in the design and development of knowledge-based agents and their functionality.

Moreover, also new concepts for high-quality services need to be designed. Such claims as "it is not natural to talk to a robot" can be understood in relation to the traditional views of the robot as a computer, an elaborated device yet operated in a similar manner as other automated devices, i.e. it is a tool to perform certain tasks. However, such views tend to overlook the other characteristics of the robot, namely its ability to move, perceive, and communicate, i.e. it is like an interactive agent which can take part in social situations. The robot thus functions in the boundary area between engineering and social science, and it is likely that with technological development of robotics the robot's social nature becomes a natural feature of task performance. The robot's autonomous characteristics define it as a boundary-crossing agent that facilitates interaction and mutual intelligibility between perspectives, while joint task performance gives rise to new concepts for designing interactive services as cooperation between humans and robots.

10.21437/AI-MHRI.2018-14

In our work, we have focussed on practical dialogue modelling that enables interactions between users and humanoid robots. The task domain is related to elder care where the robot agent is expected to act in a cooperative manner: it can advise caretakers about the normal course of actions for doing particular tasks, and also offer a friendly companion service for elderly in everyday situations.

2 Knowledge for Intelligent Agents

We explore open-domain information access systems that allow users to conduct natural interaction on varied topics. The challenge is that such interactive information agents require rich knowledge of the participants and the context in which the dialogues take place. Compared with common task-based spoken dialogue systems, our application differs from them in that (1) interactions are situated: they occur in a particular situation with a particular user in a particular context, and (2) in our scenarios, the user can either look for structured information of how to do tasks or conduct unstructured searches through digital repositories.

As discussed above, the interfaces *afford* intuitive and natural interaction which allows users to talk about their activities, emotions, and experiences, and also enables the system to understand human behavior, intentions, and awareness to support operational efficiency ([3],[8],[9]). Although it is possible to provide a sketch of the possible interactions in terms of task items as in task-based dialogue systems, it is not possible to anticipate interactions in open domain search, since they depend on the user's interest and likings rather than a task structure (cf. [7]). In micro environments, agent communication deals with partners in the immediate vicinity of the agent, and the robot agent thus needs to attend the user's interests and be aware of the user's understanding through the multimodal signals that the user emits through face, gaze and gestures. Furthermore, the robot agent needs to be equipped with knowledge of what it means to communicate. The principles behind turn-taking, topic management, reasoning and inferencing are not easy to integrate into a dialogue model, especially into neural models which are commonly trained on surface level word sequences only. However, if the robot is to be used in real-world situations such as in health-care and elder-care, education, or simply in friendly informative chatting, the robot assistant needs to provide truthful and trustworthy information in a nice, amicable manner.

The robot system is also expected to receive information through its IoT channels [11]. When communicating with smart objects in such macro environment, it is important to construct shared context for the interaction and to connect tangible devices to the intangible knowledge that humans possess about the relevant activities on these devices.

Communication in both types of environments requires knowledge, and it is essential to structure the knowledge into a systematic collection of "things", or ontologies that encode the knowledge in a hierarchical structure of concepts. Ontologies should not only represent objects and their hypernyms and hyponyms, but also actions and events in which the objects participate (event semantics). Moreover, representation formalism and its interface should be flexible and allow encoding of knowledge from different perspectives since humans have learnt the knowledge through different experience when acting and manipulating the objects. Knowledge types are discussed more in [6].

Figure 1 depicts the context for the two types of interaction envisaged: the traditional face-to-face dialogues among the agents in their immediate context (micro environment) and the IoT communication protocols that extend the context by enabling interaction with smart objects (macro environment). In macro environments, the communicating partner can be an individual or a group of objects, but the partner may also remain vague or anonymous as when using knowledge from digital repositories or social media.

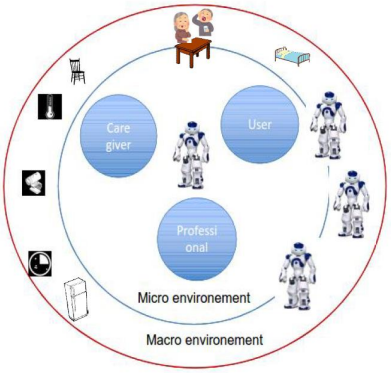

Figure 1 ***Contexts for intelligent agent systems: agent communication (micro environment) and IoT communication (macro environment).***

3 Constructive Dialogue Modelling

The interaction model incorporates some of the issues examined in Theory of Mind [12], Situated Cognition [1], and in Constructive Dialogue Modeling [3] to construct a shared context for mutual understanding and communication in a natural setting. Social interaction governs the individual's conduct and especially grounding and meaning creation processes [2] which make the world interpretable to the human agent and allow their experiences to be shared.

The interaction cycle is depicted in Figure 2. Conversational interactions are cooperative activities through which the interlocutors build a common ground (in conflict situations some level of cooperation is also needed in order to be involved in the communication in the first place and to conduct argumentation of the conflicting information). The participants are thus regarded as rational agents, engaged in cooperative activity whereby they try to achieve an underlying goal by means of exchanging new information on the topics related to the goal and their intentions. The agents keep track of the newly introduced concepts, topics, and mutually understood concepts, and coordinate their action to align themselves and to share knowledge about the task and dialogue situation.

The agents operate on the levels of Contact, Perception, Understanding and Reaction which define enablements for interaction. The two first enablements, being in contact and

perceiving communicative signals, indicate the participant's awareness of the communication: the agent pays attention to the social signals that the partner produces, their distance from each other, and that the signals are to be recognized as communicative signals. If communication is established, the agents also intend to engage in the interaction. This includes some effort to understand the partner's message: ground the concepts, establish joint goals, and build the shared context. Moreover, the agents need to plan their own reaction as a response to partner's utterance; the reaction conveys new information about the agent's intention and emotional state in the current dialogue situation.

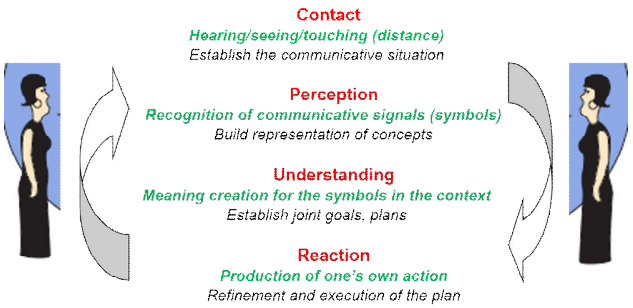

Figure 2 Constructive Dialogue Model [1].

4 System Architecture

The theoretical view of Constructive Dialogue Modelling is translated into a dialogue system, the components of which are shown in Figure 3. The primary goal has been to build a usable, integrated interaction system, but attention was also paid to creating re-usable components and methods. The dialogue model has been applied to a robot, and its operation has been presented in more detail e.g. in [5],[6],[7].

The robot system is intended to act and react to the users' responses as an "equal" partner. Signal Detection (= Contact) refers to face recognition: the robot recognizes a human face and produces greetings. ASR (Automatic Speech Recognition) (= Perception) recognizes the user utterances and via the NLP module translates them into ontology-based concepts. Topic spotting is used to find the current topic. The Interaction and Decision-making modules function as the main coordinators of the robot's communicative behavior (= Understanding), fetching knowledge of the requested topic and initiating the robot's response. The response is represented in the System Agenda as a string of words, which the TTS module speaks out to the user (= Reaction). Interaction module also keeps a record of the dialogue history, a linked list of dialogue states at a given moment including topics introduced in the interaction (what has been talked about) and the current Proposal (what task steps have been solved and what should be discussed further).

Digital Knowledge base includes task hierarchies on care-taking tasks, and they are written in json format [6]. The current implementation includes only basic task concepts, their possible risks, and details of the necessary tools, but an ontology describing concepts and their relations in the task environment is envisaged for more general language capabilities (e.g. reference resolution, lexical ambiguity resolution) and for elaborated inferencing of the actions necessary to complete the task. Two types of knowledge structures are currently used: Proposal which encodes task information (in our case: care taking tasks) and the User/System Agenda which encodes the user/robot utterance representation. The event-based semantic representation produced by a natural language parser is linked to the knowledge representation that describes what the event refers to in the environmental context, e.g. what it means to change a person's position, or provide help in dressing.

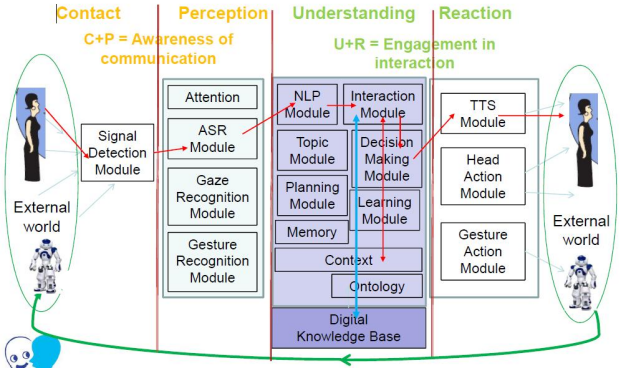

Figure 3 ***System architecture in the CDM framework, with a flow of information from speech signal travelling through some of the modules belonging to Contact, Perception, Understanding and Reaction.***

The architecture for care-taking dialogues has been implemented using Softbank's (previously Aldebaran) NAO robot platform, see [6]. This is a humanoid robot with capabilities to speak, move, and gesture. Fig. 4 shows a user interacting with the NAO robot, while examples of the first version on instruction scenarios from the care-taking domain can be found in the following links in English:
https://1drv.ms/v/s!AsdP-COrgARfgxjpPSf7Ody3QGK0
and in Japanese:
https://1drv.ms/v/s!AsdP-COrgARfgxl2C8DbAvF-gyKM

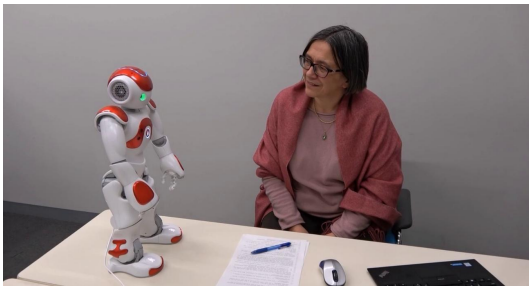

Figure 4 A user interacting with the Nao robot.

5 Discussion

This paper addressed issues related to dialogue models for social robots, aiming to fulfil requirements for natural and intuitive multimodal interaction, and taking into account human activities, rich knowledge of the world, as well as the

micro and macro environments where the agent operates. The robot dialogues aim to afford intuitive interactions in natural language and present information to the user in a collaborative manner.

A combination of various types of ontological and lexical resources in a common representational framework relates to the philosophical question of closed world knowledge and to the completeness of the knowledge included in the formalized knowledge repositories and large databases. In the context of affordable natural language robot interactions, this translates into a question of trust between the human user and the communicative robot: can the robot present information in a trustworthy manner, and if so, is the information it presents reliable.

An interesting question is also if it is possible to find ways to integrate the earlier classical top-down knowledge representation formalisms into those needed in more recent bottom-up data-driven methods. It seems possible that modularity and clean interfaces, besides suitable representations, can help to extend the knowledge needed in AI research towards natural interactive systems.

The knowledge needed for different tasks and applications is huge, and its partitioning into concepts is a matter of interpretation and in practice usually bound to the needs of the application. As the "elementary" part of the knowledge is dependent on the view-point and on the chosen level of granularity (e.g. do we talk about an elderly person or a next-door neighbor, furniture or a bed), the challenge is to enable and maintain such flexibility in the ontology. It seems necessary to explore if interdependence, collaboration, and reciprocity could indeed go beyond the limitations of the earlier hand-crafted and formalized knowledge bases, and also enable deep learning connectionists to develop techniques beyond surface level end-to-end models and include more knowledge in the process to allow more computationally tractable but also more explicable human-computer interactions.

develop techniques beyond surface level end-to-end models and include more knowledge in the process to allow more computationally tractable but also more explicable human-computer interactions.

Finally, we also pay attention to the user's personal data and individual preferences. Various levels of privacy issues are relevant for the development of talking social robots in real contexts and need to be taken into account and carefully considered when designing applications, starting from data security to privacy issues.

Open questions for discussion

The following issues are regarded as important issues to discuss when designing and implementing intelligent interactive robot agents:

1. What kind of knowledge is needed, and how to represent it?
2. How to include inferences necessary for decision making?
3. Can the structured knowledge be modelled using deep learning techniques?
4. How to modify existing information and add new information, how to explore consistency of the data, reliability and interpretability of answers?
5. How to determine appropriate responses in the rich information context for the robot?
6. How to enable the robot's attention to be guided in the interaction context?
7. How the robot's dual essence as an elaborated computer and an interactive agent influence the application development and the interaction modelling?

References

[1] Clancey W. J. (1997). *Situated cognition. On human knowledge and computer representations*. Cambridge: Cambridge University Press

[2] Harnad, S. (1990). The symbol grounding problem. *Physica D* 42: 335–346.

[3] Jokinen, K. (2009). *Constructive Dialogue Modelling – Speech Interaction with Rational Agents*. John Wiley & Sons, Chichester, UK. http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470060263.html

[4] Jokinen, K., McTear, M. (2009). Spoken Dialogue Systems. Morgan and Claypool Publishers.

[5] Jokinen, K., Nishimura, S., Fukuda, K., Nishimura, T. (2017). Dialogues with IoT Companions - Enabling human interaction with intelligent service items. Procs of the 2nd International Conference on Companion Technology (ICCT 2017), IEEE, 2017, pp. 1-3.

[6] Jokinen, K., Nishimura, S., Watanabe, K., Nishimura, T.(2018). Human-Robot Dialogues for Explaining Activities. Proceedings of IWSDS-2018, Singapore.

[7] Jokinen, K., Wilcock, G. (2013). Multimodal open-domain conversations with the Nao robot. In: Natural Interaction with Robots, Knowbots and Smartphones: Putting Spoken Dialogue Systems into Practice, pp. 213–224. Springer.

[8] Marin-Urias, L.F., Sisbot, E.A., Pandey, A.K., Tadakuma, R., Alami, R. (2009). Towards shared attention through geometric reasoning for human robot interaction. In: Humanoids 2009. The 9th IEEE-RAS International Conference on Humanoid Robots, pp. 331-336.

[9] Moratz, R., Tenbrink, T. (2008). Affordance-Based Human-Robot Interaction. *Towards Affordance-Based Robot Control*. Lecture Notes in Computer Science 4760 pp. 63 – 76

[10] Norman, D. (1988). The Design of Everyday Things. Basic Books.

[11] Smith, I. G. (Ed. 2012). The Internet of Things 2012: New Horizons. IERC-Internet of Things European Research Cluster. Halifax, U.K

[12] Wimmer, H., Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13: 103–128.