Representational Learning in Conversational Agents

Srđan Savić[1], Milan Gnjatović[1], Dragiša Mišković[1] and Branislav Borovac[1]

***Abstract*—In this paper, we report on a prototype conversational agent for natural language human-robot interaction, integrated with the assistive human-like robot MARKO. The presented prototype conversational agent is based on an extended and upgraded functional architecture encapsulating the novel functionality of representational learning in natural language human-machine interaction, primarily based on semantic categorization and associative learning of mental representations. The proposed approach is domain-independent, and in this paper it is illustrated for an interaction domain related to robot-supported therapy.**

***Index Terms*—Conversational agent, automatic domain modeling, representational learning, focus tree, human-robot interaction.**

I. Introduction and Motivation

THE idea of living and working co-existence of humans and robots, situated in dynamic human environments, requires the robots to exhibit adaptive behavior with respect to the task they execute, the current situational context, the history of interaction, etc. This adaptive behavior relates both to verbal and nonverbal interaction, indicating the necessity of an adaptive dialog management in conversational robotic agents.

In order to successfully integrate robots in human environments, it is also important to enable intelligent-like and socially-believable robot behavior, i.e. to adapt robots' behavior to established norms of social communication. Since the language, both in the spoken and written forms, is the fundamental and the most common type of expression and communication between humans, it is necessary to provide robots with the ability to participate in natural language human-robot interaction.

Besides the common functionalities underlying conversational agents, such as automatic speech recognition and text-to-speech synthesis, natural language human-robot interaction is essentially based on the adaptive dialogue management, including, but not limited to: knowledge representation, context-based interpretation of dialogue acts and decision making, learning from interaction, etc.

In this paper, we summarize selected aspects of our long-term research on the development of the human-like conversational robot MARKO [1]. We present the prototype conversational agent system. It is a cognitively-inspired dialogue system for natural language human-robot interaction, integrated with the robot that expands our previously reported systems [2, 3] by implementing a novel functionality of automatic domain modeling. The reported approach is domain-independent, but illustrated in the context of therapy of children with cerebral palsy, which is the originally intended domain of application for the assistive robot MARKO.

The structure of the paper is as follows. Section II presents some state-of-the-art solutions. In Section III, the main contribution of the paper is presented, related to the proposed approach to representational learning in human-robot interaction. In Section IV, we report on a prototype conversational agent integrated with the robot MARKO. Finally, Section V provides a brief discussion of the presented work and a conclusion.

II. Related work

Modeling of an interaction domain requires an appropriate form for knowledge representation as well as an ability to acquire new information from interaction, i.e. to learn novel mental representations. Both of these processes require some sort of memory system, and they are particularly related to its semantic aspect. Three main methodological approaches to cognitive modeling can be distinguished: (i) statistical, (ii) symbolic and (iii) connectionist.

Some state-of-the-art statistical models of semantic memory include LSA model (Latent Semantic Analysis) [4], HAL (Hyperspace Analogue to Language) [5] and BEAGLE model (Bound Encoding of the Aggregate Language Environment model) [6]. These models belong to the class of so-called distributional models, since they are grounded in the distributional hypothesis [7]. Common point for these models is the fact that they are related to statistical semantics and use the redundancy in language to learn semantic representations.

Some classical models of semantic representations include hierarchical network model [8] and spreading-activation model [9]. It is important to mention hybrid computational cognitive architecture ACT-R (Adaptive Control of Thought-Rational) [10]. This architecture has been applied in several

[1] Srđan Savić is with the Faculty of Technical Sciences, University of Novi Sad, 6 Trg Dositeja Obradovića, 21000 Novi Sad, Serbia (e-mail: savics@ uns.ac.rs).

[1] Milan Gnjatović is with the Faculty of Technical Sciences, University of Novi Sad, 6 Trg Dositeja Obradovića, 21000 Novi Sad, Serbia (e-mail: milangnjatovic@ uns.ac.rs).

[1] Dragiša Mišković is with the Faculty of Technical Sciences, University of Novi Sad, 6 Trg Dositeja Obradovića, 21000 Novi Sad, Serbia (e-mail: dragisa@ uns.ac.rs).

[1] Branislav Borovac is with the Faculty of Technical Sciences, University of Novi Sad, 6 Trg Dositeja Obradovića, 21000 Novi Sad, Serbia (e-mail: borovac@ uns.ac.rs).

domains, relevant for this paper, like: semantic categorization [11] and associative learning [12].

Regarding the class of connectionist models it is important to mention fusion adaptive resonance theory for multimemory learning [13]. It is a neural architecture which provides a unified framework for representation and learning of various types of semantic knowledge and addresses the relation between semantic, episodic and procedural memory modules. In paper [14] another model, based on fusion adaptive resonance theory, was used to provide unified approach to a set of distinct learning paradigms.

Another approach to problem of multi-domain dialogue management, based on the hierarchical reinforcement learning is presented in [15], while a method for multi-domain dialogue policy deep reinforcement learning—termed NDQN is presented in [16], which applies to an information-seeking spoken dialogue system in the domains of restaurants and hotels.

III. Representational Learning in Human-Machine Interaction

Appropriate modelling of a dialogue domain is considered fundamental for successful human-robot dialogue management [17, 18]. In addition, modelling of a wider interaction context is essential for natural and long-term interaction. The conversational agent reported in this paper is built upon the focus tree model, which represents the core of the formerly introduced cognitively-inspired approach to meaning representation in human-machine interaction [2]. Unlike the other well-known approaches to meaning representation, like frames, semantic networks, etc. [17], the focus tree model addresses the problem of capturing the meaning of the spontaneously produced utterances, without a preset grammar and syntactic constraints.

In this section, we describe an extension of the focus tree model. Namely, a focus tree is a hierarchical structure representing a dialog domain, whose nodes represent semantic entities, i.e. concepts, and directed edges represent relations between them. Directed top-down paths constitute mental representations [3], and – among them – directed paths that connect the root node with terminal nodes represent basic mental representations, called propositions. Thus, a knowledge related to a dialog domain is encoded in the focus tree, both through its topology and semantic content.

In the early version of the model, the knowledge of a dialog domain was hand-coded [2, 3], and later a corpus-based approach to automatic corpus-driven domain modelling was presented in [19]. In contrast to this, a symbolic approach to automatic domain modelling was proposed in [20], as a part of a cognitively-inspired computational memory model. This approach to representational learning provides the ability of incremental on-line learning of novel mental representations and dynamic modification of the existing domain model. The point of departure for this learning algorithm is a set of propositions, which only partially describe a set of dialogue domains. Input set contains propositions, related to different domains, while the number of these domains is arbitrary and initially unknown to a conversational agent. The approach is based on two mechanisms, the first one of semantic categorization of semantic entities, and the second of associative learning of novel propositions. Thus, given an input set of propositions, the proposed learning algorithm performs the following tasks:

- semantic classification of propositions, according to dialogue domains they belong to,
- associative inferring of missing propositions, in order to complete the knowledge of the underlying domains.

Therefore, the system identifies the number of different domains represented in the input set, and performs learning by analogy to infer novel propositions which describe the underlying domains [20]. The proposed algorithm is symbolic, which makes it computationally and analytically tractable.

In this paper, we propose an integration of natural language understanding model and a dialogue management model, both of which are based on the focus tree model, with additional module based on novel approach to representational learning, in order to support the process of automatic domain modeling in human-machine interaction.

IV. Prototype Conversational Agent

This section reports on a prototype conversational agent integrated with the assistive human-like robot MARKO, cf. Fig. 2. The prototype is based on a conversational agent which includes the novel module for representational learning. The robot MARKO is designed as an assistive tool for therapy of children with cerebral palsy [1]. There are several robot-supported therapeutic scenarios, primarily related to gross and fine motor skills exercises, where the conversational agent, integrated with the robot, is aimed at increasing the child's motivation by means of engagement in a three-party natural language interaction between the therapist, the child, and the robot [21, 22, 23]. In the following text, we briefly describe the interaction domains related to the therapy, and present the architecture of the prototype conversational agent.

Fig. 1. Human-like conversational robot MARKO.

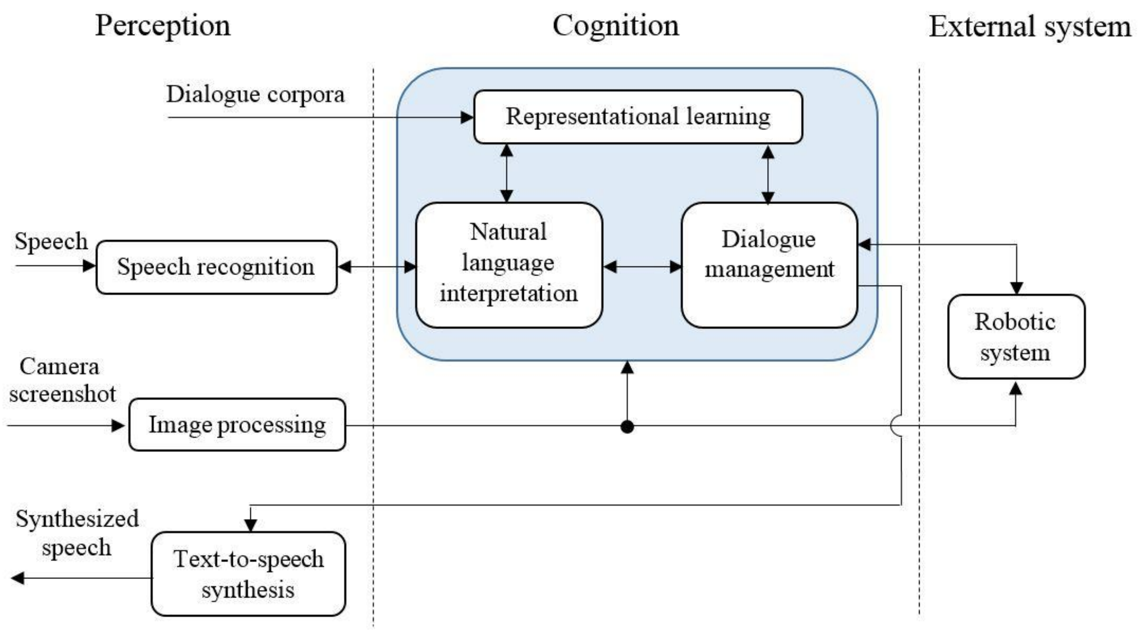

Fig. 2. Architecture of the prototype conversational agent.

A. Dedicated Domains of Interaction

The first interaction domain relates to gross motor skill exercises and assumes bidirectional speech-based interaction between the therapist and the robot, and a unidirectional interaction between the robot and the child, in which the robot addresses the child verbally [21]. The therapist requires from the robot to perform an exercise, and the robot normally responds with a non-verbal action, in the form of certain movement to demonstrate the required exercise, and an accompanying verbal dialogue act. The robot may also require additional information from the therapist, if needed, in order to interpret the perceived command. Also, the robot may verbally instruct the child to perform an exercise and it may encourage the child with appropriate verbal acts.

The second interaction domain relates to fine motor skill exercises, and contains a set of wooden three-dimensional objects which differ in several modalities: shape, color, thickness and size. The robot may move an object in four different spatial directions (leftward, rightward, upward, and downward) or it may point to an object. Similarly as in the first interaction domain, all interacting participants share the verbal and spatial contexts.

B. Architecture of the Conversational Agent

The architecture of the prototypical conversational agent is constituted from modules that can be classified in three distinctive categories: perception, cognition and an external hardware system, respectively (cf. Fig. 2). In the following text, each module of the architecture, is briefly described:

(i) Perception:

- *Speech recognition module* – is based on a novel hybrid methodological approach to context-dependent speech recognition [24]. This approach accounts for a dialogue context, unlike the traditional statistical approaches which only account for the context at the level of sentences or phonemes. Speech recognition is implemented for the Serbian language, and includes a vocabulary of approximately 5000 words.
- *Image processing module* – includes two pairs of cameras. The first pair of cameras is mounted on a fixed stand beside the table situated in the working space of the robot, and is used for stereovision. The functionality of this submodule is to recognize different geometrical objects placed on the table, and is implemented by using the open source library for real time computer vision *openCV* [25]. Another pair of cameras is placed in the eyes of the robot. These cameras are used for the tracking of the user's face, so that the user is given an impression that the robot keeps an eye contact with her, which contributes to more socially-believable interaction .
- *Speech synthesis module* – performs text-to-speech synthesis of the robot's dialogue act in the Serbian language [26].

(ii) Cognition:

- *Representational learning module* – is based on the proposed symbolic approach to automatic domain modelling in task-oriented human-machine interaction. It encapsulates the functionalities of semantic categorization and associative learning [20].
- *Natural language interpretation module* – extracts semantic information from the spontaneously produced linguistic input and interprets the user's utterance. The approach assumes no preset grammar or any constraints on the syntactic structure of the user's utterance [2].
- *Dialogue management module* – Based on a dialogue context model and a dialogue control model (i.e., dia-

logue strategies), this module manages the interaction between the user and the robot [2].

(iii) External system:

In the presented prototype, the human-like robot MARKO [1] is applied as the external system. It is a robotic platform with an anthropomorphic kinematic structure. It has 33 degrees of freedom, excluding the hands, which are still under development [27, 28]. The kinematic structure and mechanical design of the robot were conditioned, to a large extent, by the specification requirements related to the therapy. The mechanical design of the robot provides sufficient manipulability and working space to cover the scope of relevant therapeutic exercises. The robot has a user-friendly visual appearance whose design was based on the psychological research conducted on a group of children [29]. The robot MARKO is equipped with different types of sensors (incremental and multi-turn absolute encoders, 3-axis force sensors in fingertips, cameras and microphone) to provide the robot proprioception and hardware support for a conversational agent. For more details about the technical implementations of the proposed architecture, we refer the reader to [30].

V. Discussion and Conclusion

This paper provides a brief summary of the results of the research conducted within a project aimed at developing a conversational human-like robot as an assistive technology for therapy of children with developmental disorders. A prototype conversational agent, integrated with a custom-designed human-like robot platform, is presented. The presented conversational agent prototype is based on an extended and upgraded functional architecture encapsulating the novel functionality of representational learning in natural language human-machine interaction, primarily based on semantic categorization and associative learning. This functionality is based on a symbolic approach to automatic domain modeling in task-oriented human-machine interaction. The proposed approach is scalable, computationally and analytically tractable, and domain-independent. In this paper, it was illustrated for a specific interaction domain.

Acknowledgment

The presented study was funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia (research grants III44008 and TR32035), and by the intergovernmental network EUREKA (research grant E!9944). The responsibility for the content of this article lies with the authors.

References

[1] B. Borovac, M. Gnjatović, S. Savić, M. Raković, M. Nikolić, “Human-like Robot MARKO in the Rehabilitation of Children with Cerebral Palsy,” in *New Trends in Medical and Service Robots. Mechanisms and Machine Science*, Bleuler H., Bouri M., Mondada F., Pisla D., Rodić A., Helmer P. (eds), vol 38. Springer, Cham, 2016, pp. 191-203.

[2] M. Gnjatović and V. Delić, “Cognitively-inspired representational approach to meaning in machine dialogue,” *Knowledge-Based Systems*, 71, pp. 25-33, 2014.

[3] M. Gnjatović, M. Janev and V. Delić, “Focus tree: modeling attentional information in task-oriented human–machine interaction,” *Appl. Intellig.* 37 (3), pp. 305–320, 2012.

[4] T. Landauer and S. Dumais, "A solution to Plato’s problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge," Psychological Review, vol. 104, no. 2, p. 211–240, 1997.

[5] K. Lund and C. Burgess, "Producing high-dimensional semantic spaces from lexical co-occurrence," Behavioral Research Meth-ods, Instrumentation, and Computers, vol. 28, no. 2, p. 203–208, 1996.

[6] M. N. Jones and D. J. K. Mewhort, "Representing word meaning and order information in a composite holographic lexicon," Psycho-logical Review, vol. 114, no. 1, p. 1–37, 2007.

[7] Z. Harris, Distributional structure. Word, 10(23): 146-162, 1954.

[8] M. R. Quillian, "Semantic memory," in Semantic Information Processing, M. U. Minsky, Ed., Cambridge, MA: MIT Press, 1968.

[9] A. M. Collins and E. F. Loftus, "A spreading activation theory of semantic memory," Psychological Review, vol. 82, no. 6, pp. 47- 428, 1975.

[10] J. R. Anderson, How Can the Human Mind Occur in the Physical Universe?. Oxford University Press, 2007.

[11] M. Rutledge-Taylor, C. Lebiere, R. Thomson, J. Staszewski and J. R. Anderson, A Comparison of Rule-Based versus Exemplar-Based Categorization Using the ACT-R Architecture, Proceedings of the 21st Annual Behavior Representation in Modeling and Simulation Conference, pp. 44-50, 2012.

[12] R. Thomson, A. Pyke, J. G. Trafton and L. M. Hiatt, An Account of Associative Learning in Memory Recall. Proceedings of the 37th Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society, pp. 2386-91, 2015.

[13] W. Wenwen, T. Ah-Hwee, T. Loo-Nin, Semantic Memory Modeling and Memory in Learning Agents, IEEE Transactions on System, Man and Cybernetics: Systems, vol. 47, no. 11, pp. 2882-2895, 2017.

[14] A. H. Tan, G. A. Carpenter, S. Grossberg, Intelligence Through Interaction: Towards a Unified Theory for Learning. In: Liu D., Fei S., Hou ZG., Zhang H., Sun C. (eds) Advances in Neural Networks – ISNN 2007. ISNN 2007. Lecture Notes in Computer Science, vol 4491. Springer, Berlin, Heidelberg, 2007.

[15] P. Budzianowski, S. Ultes, P. Su, N. Mrksic, T. Wen, I. Casanueva, L. M. Rojas-Barahona and M. Gasic, Sub-domain Modelling for Dialogue Management with Hierarchical Reinforcement Learning. SIGDIAL Conference, 2017.

[16] H. Cuayáhuitl, S. Yu, A. Williamson and J. Carse, Deep Reinforcement Learning for Multi-Domain Dialogue Systems, NIPS Workshop on Deep Reinforcement Learning, arXiv preprint arXiv:1611.08675, 2016.

[17] D. Jurafsky and J. Martin, *Speech And Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, 2nd edition, Upper Saddle River, NJ, USA, Prentice Hall, 2009.

[18] K. Jokinen and M. McTear, “Spoken Dialogue Systems,” in *Synthesis Lectures on Human Language Technologies*, Morgan and Claypool, 2(1), pp. 1-151, 2009.

[19] M. Gnjatović and V. Delić, “Electrophysiologically inspired evaluation of dialogue act complexity,” Proc. of the 4th IEEE International Conference on Cognitive Infocommunications, CogInfoCom, Budapest, Hungary, pp. 167-72, 2013.

[20] S. Savić, M. Gnjatović, D. Mišković, J. Tasevski and N. Maček, “Cognitively-Inspired Symbolic Framework for Knowledge Representation,” Proc. of the 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Debrecen, Hungary, September 2017, pp. 315-320, 2017.

[21] M. Gnjatović, “Therapist-Centered Design of a Robot’s Dialogue Behavior,” *Cognitive Computation*, 6(4), pp. 775-788, 2014.

[22] M. Gnjatović, J. Tasevski, D. Mišković, S. Savić, B. Borovac, A. Mikov, R. Krasnik, “Pilot Corpus of Child-Robot Interaction in Therapeutic Setting,” 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Debrecen, IEEE, pp. 253-257, 11-14 September, 2017.

[23] M. Gnjatović, “The Child, the Therapist, and the Robot: Adaptive Dialogue Management in Three-Party Interaction,” *COST Workshop on Social Robotics*, The Future Concept and Reality of Social Robotics: Challenges, Perception and Applications - The Role of Social Robotics in Current and Future Society, Brussels, 10-13 June 2013, *Invited*.

[24] D. Mišković, M. Gnjatović, P. Štrbac, B. Trenkić, J. Nikša and V. Delić "Hybrid methodological approach to context-dependent speech recognition," *International Journal of Advanced Robotic Systems*, 14(1), 12 pages, no pagination, 2017.

[25] J. Tasevski, M. Nikolić and D. Mišković, "Integration of an Industrial Robot with the Systems for Image and Voice Recognition," *Serbian journal of electrical engineering*, Vol. 10, No. 1, pp. 219-230, 2013.

[26] V. Delić, D. Pekar, R. Obradović, N. Jakovljević and D. Misković, "A Review of AlfaNum Continuous Automatic Speech Recognition System," XII International Conference "Speech and Computer", Moscow, Russia, 15 – 18 Oct. 2007.

[27] S. Savić, M. Raković, M. Penčić, M. Nikolić, S. Dudić and B. Borovac, "Design of an Underactuated Adaptive Robotic Hand with Force Sensing," 3rd International Conference on Electrical, Electronic and Computing Engineering (IcETRAN), ETRAN Society, Zlatibor, 13-16 June, 2016.

[28] M. Raković, G. Anil, Ž. Mihajlović, S. Savić, S. Naik, B. Borovac and A. Gottscheber, "Fuzzy Position-Velocity control of Underactuated Finger of FTN Robot Hand", *Journal of Intelligent and Fuzzy Systems*, ISSN: 1064-1246, Vol. preprint, No. preprint, 2018.

[29] M. Oros, M. Nikolić, B. Borovac and I. Jerković, "Children's preference of appearance and parents' attitudes towards assistive robots," 14th IEEE-RAS International Conference on Humanoid Robots (Humanoids) 2014, pp. 360-365, 18-20 Nov. 2014.

[30] M. Gnjatović, D. Mišković, S. Savić, B. Borovac, N. Maček and B. Trenkić, "A Novel Modular Architecture for Conversational Agents," 4th International Conference on Electrical, Electronic and Computing Engineering IcETRAN, Kladovo, 5-8 June, 2017.