Humanoid Robots and Spoken Dialog Systems for Brief Health Interventions

Saminda Abeyruwan
University of Miami
Dept. of Computer Science
Coral Gables, FL 33146
saminda@cs.miami.edu

Ramesh Baral
Florida Int'l University
Comp. & Info Science
Miami, FL 33199
rbara012@cis.fiu.edu

Ugan Yasavur
Florida Int'l University
Comp. & Info Science
Miami, FL 33199
uyasa001@cis.fiu.edu

Christine Lisetti
Florida Int'l University
Comp. & Info Science
Miami, FL 33199
lisetti@cis.fiu.edu

Ubbo Visser
University of Miami
Dept. of Computer Science
Coral Gables, FL 33146
visser@cs.miami.edu

Abstract

We combined a spoken dialog system that we developed to deliver brief health interventions with the fully autonomous humanoid robot (NAO). The dialog system is based on a framework facilitating Markov decision processes (MDP). It is optimized using reinforcement learning (RL) algorithms with data we collected from real user interactions. The system begins to learn optimal dialog strategies for initiative selection and for the type of confirmations that it uses during the interaction. The health intervention, delivered by a 3D character instead of the NAO, has already been evaluated, with positive results in terms of task completion, ease of use, and future intention to use the system. The current spoken dialog system for the humanoid robot is a novelty and exists so far as a proof of concept.[1]

Motivation and background

Latest developments spoken dialog systems (SDS) show complementary progress for the development of embodied conversational agents (ECA) (Yasavur and Lisetti 2014). Yet, very little of this progress has been used for autonomous humanoid robots, which have also recently demonstrated the ability to serve in health interventions. The humanoid robot NAO specifically has been used in a variety of applications in the health domain (e.g., (Greczek, Atrash, and Matarić 2013)) including user studies which demonstrate the usefulness of the NAO platform.

Dahl and Boulos (2013) gave a recent overview of how robots are used in healthcare. Apart from the surgical robots that are tele-operated by a human doctor, robots support the daily work in hospitals, mainly in logistics (e.g., Atheon TUG platform (Bloss 2011) or HelpMate (Evans and Krishnamurthy 1998)). Several other studies have looked at different autism spectrum disorders. The main goal is to provide therapy with the help of an intelligent robotic agent that improves both social and communication skills of the children involved. One example is the KASPAR robot (Robins et al. 2012). Its creators use robot-based play scenarios that can be tailored for types of disability and skill areas that need to be motivated. Another example has been created by Bekele and colleagues. They focus on communication behaviors, in particular head-tracking to manifest the robots engagement in the on-going interaction (Bekele et al. 2013).

There are also several studies involving the humanoid robot NAO. Csala and colleagues, for example, studied the effectiveness of a tele-operated NAO humanoid robot in improving the wellbeing of children having undergone marrow-transplants. Although there is no conversation with the user involved, the study demonstrated that the NAO robot is well suited for this task due to its small size and robustness (Csala, Nemeth, and Zainko 2012). Also, the study identified personalization as a key requirement for success. The NAO robot has also been used by Belpaeme and colleagues who used the robot to both entertain and educate children suffering from diabetes in a hospital environment (Belpaeme et al. 2012). This work is interesting to our approach as it is focused on providing high levels of robot autonomy using a natural language interface and a long-term memory structure that allowed children to develop a personal relationship with the robot. Both studies by Csala and colleagues and Belpaeme and colleague took place in hospitals, and both efforts were greatly appreciated by the children involved.

In this article, we discuss a new mode of delivery - a spoken dialog system coupled with a NAO robot - for health interventions that can help people adopt healthy lifestyles. The dialog system can deliver brief health interventions (BI), which are short, well structured, one-on-one counseling sessions, focused on specific aspects of problematic lifestyle behavior (e.g. overeating, poor diet, heavy drinking). BIs, which are top ranked out of 87 treatment styles in terms of efficiency (Miller and Wilbourne 2002) and which can be

[1]We plan to give a demo if our contribution would be accepted.

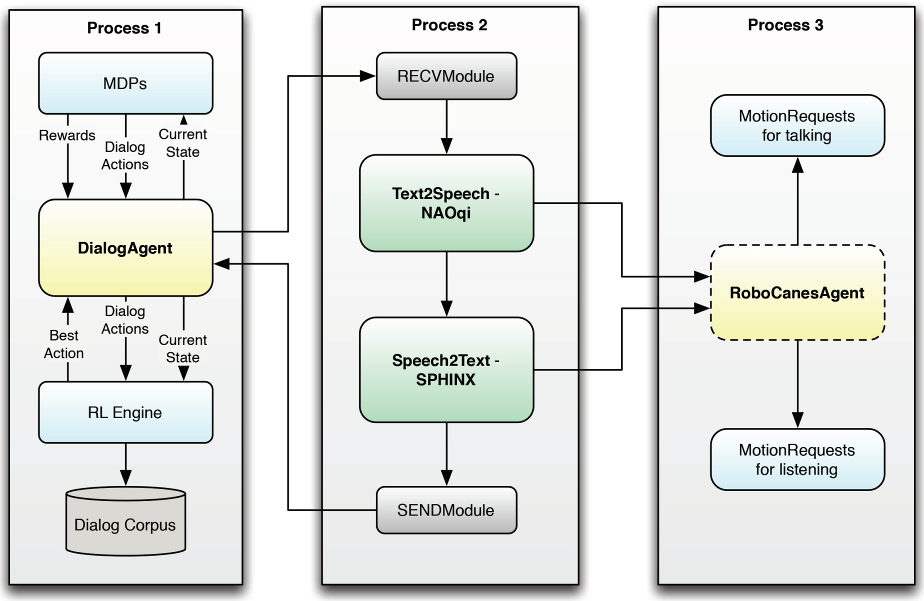

Figure 1: System Architecture: process 1 contains the dialog manager, process 2 delivers speech from the robot and the recognition of the human, whereas process 3 is responsible for motions that accompany the robot's speech and listening time.

delivered in 3-5 minutes (Moyer et al. 2002), assess a person's problem behavior, and provide advice about ways to eliminate it.

Approach and implementation

Our approach is based on two system designs that have been invented and developed in two different research groups and independently. The FIU Affective Social Computing group has developed a dialog manager that is integrated in a system using a virtual character. Our system utilizes the users natural speech as one input parameter and also uses gaze and non-verbal cues for the classification of the users state. The dialog manager operates with five interconnected MDPs representing each step or phase of the behavior intervention (Yasavur and Lisetti 2014). The dialog manager uses domain specific dialog phrases, a tailored grammar for this case and a RL engine that learns over time how to address/answer to the user. It has been successfully tested with 87 users for the alcohol domain.

The University of Miami's autonomous robots research group develops methods for both simulated and physical agents that act in dynamic, adversarial, and real-time environments. The group uses the RoboCup domain as an area of investigation. One focus is on appropriate consideration of agent's actions/behavior in order to make own decisions (for the own team agents). The group has developed a fully implemented system (RoboCanes software) using six NAO robots that not only move in the mentioned environment on their own, but also have to consider team communication (e.g. for global object positions such as the ball or an opponent agent) and cooperation. The software is stable and robust and is used as part of the backbone for this research paper.

We ported the dialog manager on one of our NAO robots. The software runs as a single process. The dialog manager is connected to another process on the robot that makes use of CMU's Sphinx speech recognizer system. We use the pocketsphinx 0.8 system, which is a flexible hidden Markov model-based speech recognition system (Huggins-Daines et al. 2006). We also use the Sphinx 5.0 english (US) standard language model without training or domain specific grammars. The original dialog system described in (Yasavur and Lisetti 2014) has a language model and grammars, which we plan to adapt for the NAO robot in our next system version. A third process is the RoboCanes software that can be connected to the dialog. Figure 1 gives an overview of the architecture and the three processes. Process 1 runs as a server that is connected to process 2 via a TCP/IP connection. The robot receives a greeting mechanism at the beginning of a conversation that can be altered when starting again. The NAO robot then uses NAOqi for the Text-To-Speech generation, although we have also implemented and tested other systems such as Festival (Taylor, Black, and Caley 1998) or eSpeak (eSpeak Dev Team 2014), which would also work. Process 2 then makes use of the RoboCanes framework (the RoboCup agent) which is responsible for robot motions and also for the image processing and the audio feedback (no feedback from cameras at the moment yet). The robot then uses pocketsphinx, turns the audio signal into text and sends it back to the dialog manager. The timing between processes 1 and 2 are based on turns, process 3 runs with 100 Hz for the joint requests and 30 Hz for both cameras.

Conclusions

Although currently at the prototype stage, we anticipate that the NAO robot will become a very likable and effective mode of delivery for brief interventions for target behaviors such as poor diet, overeating, or lack of exercise, among others. The appeal of the NAO to children (Belpaeme et al. 2012) makes it particularly suitable to become a child's favorite health coach, say, to discuss eating more fruits and vegetables on a daily basis.

References

Bekele, E. T.; Lahiri, U.; Swanson, A. R.; Crittendon, J. A.; Warren, Z. E.; and Sarkar, N. 2013. A step towards developing adaptive robot-mediated intervention architecture (aria) for children with autism. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on* 21(2):289–299.

Belpaeme, T.; Baxter, P. E.; Read, R.; Wood, R.; Cuayáhuitl, H.; Kiefer, B.; Racioppa, S.; Kruijff-Korbayová, I.; Athanasopoulos, G.; Enescu, V.; et al. 2012. Multimodal child-robot interaction: Building social bonds. *Journal of Human-Robot Interaction* 1(2):33–53.

Bloss, R. 2011. Mobile hospital robots cure numerous logistic needs. *Industrial Robot: An International Journal* 38(6):567–571.

Csala, E.; Nemeth, G.; and Zainko, C. 2012. Application of the nao humanoid robot in the treatment of marrow-transplanted children. In *Cognitive Infocommunications (CogInfoCom), 2012 IEEE 3rd International Conference on*, 655–659.

Dahl, T. S., and Boulos, M. N. K. 2013. Robots in health and social care: A complementary technology to home care and telehealthcare? *Robotics* 3(1):1–21.

eSpeak Dev Team. 2014. eSpeak text to speech. Webpage: http://espeak.sourceforge.net.

Evans, J. M., and Krishnamurthy, B. 1998. Helpmate®, the trackless robotic courier: A perspective on the development of a commercial autonomous mobile robot. In *Autonomous Robotic Systems*. Springer. 182–210.

Greczek, J.; Atrash, A.; and Matarić, M. 2013. A computational model of graded cueing: Robots encouraging behavior change. In Stephanidis, C., ed., *HCI International 2013 - Posters' Extended Abstracts*, volume 374 of *Communications in Computer and Information Science*. Springer Berlin Heidelberg. 582–586.

Huggins-Daines, D.; Kumar, M.; Chan, A.; Black, A. W.; Ravishankar, M.; and Rudnicky, A. I. 2006. Pocketsphinx: A free, real-time continuous speech recognition system for hand-held devices. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, volume 1, I–I. IEEE.

Miller, W. R., and Wilbourne, P. L. 2002. Mesa grande: a methodological analysis of clinical trials of treatments for alcohol use disorders. *Addiction* 97(3):265–277.

Moyer, A.; Finney, J. W.; Swearingen, C. E.; and Vergun, P. 2002. Brief interventions for alcohol problems: a meta-analytic review of controlled investigations in treatment-seeking and non-treatment-seeking populations. *Addiction* 97(3):279–292.

Robins, B.; Dautenhahn, K.; Ferrari, E.; Kronreif, G.; Prazak-Aram, B.; Marti, P.; Iacono, I.; Gelderblom, G. J.; Bernd, T.; Caprino, F.; et al. 2012. Scenarios of robot-assisted play for children with cognitive and physical disabilities. *Interaction Studies* 13(2):189–234.

Taylor, P.; Black, A. W.; and Caley, R. 1998. The architecture of the festival speech synthesis system. In *Proc. Third ESCA/COCOSDA Workshop on Speech Synthesis*, 147 – 151. International Speech Communication Association.

Yasavur, U., and Lisetti, C. 2014. Intelligent virtual agents and spoken dialog systems come together to deliver behavior change health interventions. *Journal on Multimodal User Interfaces, in press*.