A Platform for Human-Robot Dialog Systems Research

Rodney D. Nielsen[1], Richard Voyles[2], Daniel Bolanos[3], Mohammad Mahoor[2], Wilson Pace[4], Katie Siek[1], and Wayne Ward[3]
[1] CU Boulder, [2] University of Denver, [3] Boulder Language Technologies, [4] CU Denver Anschutz Medical Campus
Rodney.Nielsen@Colorado.edu

Abstract

We present our position on developing an open source platform for human-robot dialog research in home healthcare. We discuss the need for perceptive, emotive, spoken-dialog robotic companions, describe an architecture for this research, and call on others to participate in making it a successful community resource.

Introduction

The overall, long-term goal of our research is to develop a fully animated robotic companion for home healthcare that possesses the ability to engage humans in a way that is unobtrusive and suspends disbelief so that the humans will act naturally in their environment. This will enable the companion robot to monitor patients for signs of physical and emotional deterioration for the purpose of notifying a remote caregiver, to provide reminders and even life coaching to extend independence and quality of life.

The core of this technology is a cloud of mobile, perceptive, emotive, animate, spoken dialog agents (robotic companions) that can understand specific events via image recognition, spoken dialog, and other sensory input and that are loosely connected via the Internet to enable them to learn from their collective experiences and harness the power of others. These companionbots will enable research into proactive dialog with patients on aspects of health and daily functions, as well as facilitate motivation, training, and education, as seen in the following example.

Training the Depressed to Avoid a Pessimistic Explanatory Style: Most of the 21 million Americans suffering from major depression receive infrequent therapy, at best, based on imperfect recollection of a few of the prior week's incidents. Studies indicate that Cognitive Behavioral Therapy (CBT) is as effective as medication in treating depression and may be more likely to lead to sustained improvement. Forms of CBT have shown positive effects when delivered via interactive voice response and computer systems (Selmi et al. 1990). These systems enable interaction with patients who might otherwise not have access due to financial means, time, social stigma, location, etc. These systems enable broad access, but they follow prespecified scenarios, cannot react to real-life situations, and require much patient initiative.

A therapeutic emotive agent that *proactively* engages patients in relevant dialog would overcome these issues and augment human therapy to facilitate interactions that would otherwise not be possible. This dialog could complement existing care via more frequent interactions and point-of-need delivery. Such companionbots would result in a means of capturing a wealth of information regarding depression, which could lead to improved treatments.

The following is a concrete scenario we will use to motivate our platform for home healthcare and related STEM research. The scenario presents how we envision the companion will train patients with depression to not think or express pessimistic thoughts in a personal, permanent and pervasive way. Ruth is a patient being treated for depression. She has just dropped her glass and spilled water on herself and the floor. Daisy is her companion robot.

Ruth: Oh for crying out loud! Why do I (*pers.*) always (*perm.*) have to drop everything (*perv.*) all the time (*perm.*)?

Daisy: Oh that's a shame, but Ruth, I haven't seen you drop a glass (*not perv.*) for a long time (*not perm.*). And I've never seen you drop plates, pots or pans (*not perv.*). Everybody (*not pers.*) drops something (*not perv.*) from time to time (*not perm.*). You shouldn't be so hard on yourself.

In order to achieve this, Daisy must recognize dishes, falling events, linguistic indicators of personal, permanent, and pervasive speech and be able to rephrase such speech using more positive patterns. The technology exists in computer vision and NLP to detect these events and in HCI to communicate in natural ways. Therefore, the time is right to integrate these technologies into a companion robot that enables research investigating such healthcare scenarios.

Research Platform

The proposed platform will enable a range of research in human language technologies, image processing, ML, robotics, and other areas. Due to the lack of such a platform, there has been relatively little prior work to address the technical requirements of home healthcare robots.

It is our long-term goal to develop companion robots that encourage humans to form an emotional attachment

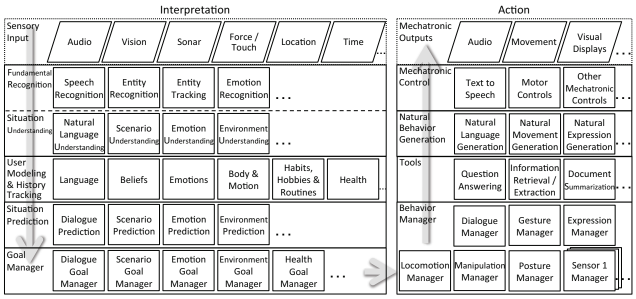

Figure 1: Companionbot Architecture

(Breazeal 2003). This affective quality is critical to home-care research, especially when the companionbot must build enough rapport to engage the participant in natural conversation. Another unique feature of the platform proposed here is that the companionbots will proactively engage the participant in dialog at opportune times, rather than passively waiting for the human to initiate interaction (Brisben, Lockerd, and Lathan 2004).

We plan to develop an open source platform that provides a solid foundation for research, while allowing researchers to easily replace any component desired. We chose the Nao, a well-designed commercial robot, as our pilot hardware platform, but also intend to test on other robots. The goal of our initial project is to develop a software platform that enables more natural human-robot interaction and includes the foundations for emotion recognition and expression, image and event recognition, speech recognition and synthesis, natural language understanding, natural spoken dialog, and improved user and environment modeling to facilitate these goals.

Key Platform Components

The Architecture of our companionbots is shown in Figure 1. One of the key differences relative to prior work is that this architecture, in general, assumes that a component in a given layer can be affected by *all* of the components in the preceding and its own layer, and occasionally by many of the components in the next layer. For example, the speech recognition component will fuse information from all of the sensory inputs in making its initial judgments. Then it might revise its final output based on the output of other fundamental recognition engines in the same layer of the architecture and based on the deeper understanding provided by components in the subsequent layer. Each layer in the architecture includes a primary component, which helps control the sequencing and interaction of the components in that layer and synthesizes their collective outputs.

The left side of Figure 1 represents the interpretation of the sensory inputs, the recording of that interpretation, and in the bottom layer, adjustments to the robot's high-level goals as a result of the interpretation. The right side of the figure depicts the processes involved in the robot's response to those sensory inputs.

Fundamental Recognition components identify and classify basic concepts based on sensor inputs for audio, vision, sonar, force, etc. This includes multimodal speech, gesture, sound, speaker, face, person, object, and emotion recognition. It also includes tracking of moving people and objects.

Situation Understanding components provide a semantic interpretation of the output of the recognition layer.

User Modeling and History Tracking components capture information related to the person's health or pertinent to assessing their capacity for activities of daily living.

Situation Prediction components will utilize the models built and maintained in the previous layer along with the situation understanding from the layer before that to infer the likely consequences or outcomes given the current situation.

Goal Manager components employ the user model & history, situation understanding and predictions produced by the previous three layers of the architecture, along with general longer-term health and wellness goals specified by the person or their healthcare provider, to revise the high level goals of the companionbot.

Behavior Manager components take the immediate goals and compute specific behaviors to achieve them.

Tools (e.g., a question answering system) will be utilized by the behavior managers to support their goals.

Natural Behavior Generation components will consume the general behavior descriptions and output of the preceding tools and will generate detailed, but hardware-independent, robotic behaviors such as dialog, gestures, facial animations, and locomotion.

Mechatronic Control components convert these platform-independent detailed descriptions into the input required by the hardware controllers.

Summary

The proposed platform will integrate language, vision, emotion, and other technologies to proactively recognize and dialog with patients on health issues such as depression. The project team is extremely multidisciplinary which will help ensure the platform is well suited for a diverse group of future researchers, not only in home healthcare robotics, but also in the numerous foundational STEM fields. The positive impact of this unique infrastructure on capabilities to conduct leading edge research is expected to be significant. How significant depends on ensuring the platform is open and comprehensive enough to provide the functionality the community needs. Making this platform a reality for all and an effective research platform will be greatly facilitated by broad community participation – by *your* participation.

References

Breazeal, C. 2003. Emotion and sociable humanoid robots. *International Journal of Human Computer Interaction* 59:119–155.

Brisben, A. J.; Lockerd, A. D.; and Lathan, C. 2004. Design evolution of an interactive robot for therapy. *Telemedicine Journal and e-Health* 10:252–259.

Selmi, P. M.; Klein, M. H.; Greist, J. H.; and et al. 1990. Computer-administered CBT for depression. *American Journal of Psychiatry* 147:51–56.