Reading 2

Discussion on Modern Conversational Agents

Subscribed

- Can we use actor voices to immediately convey backstory and evolving story line?

Display mode Display replies flat, with oldest first Display replies flat, with newest first Display replies in threaded form Display replies in nested form

Discussion on Modern Conversational Agents

by Bo Cao - Sunday, January 29, 2017, 6:25 PM

We have made many advances in natural language processing (NLP) since the "Believable Robot Characters" article was published. However, most modern conversational agents (Siri, Alexa, Google Home) do not make much use of either embodiment or the framework of "dramatic structure with rich backstory and evolving story line, verbal and nonverbal social behaviors, and believable culturally specific characters" (or maybe you'd argue that they do?). On the other hand, the Jibo robot represents a project that does strive to take advantage of such elements. How important do you think these elements are? Do you think the Jibo robot will be successful?

Me:

In my opinion, apart from merely verbal and some optical responses (like various light colors on the edges of Alexa, the graphs on Google Home indicating she's speaking, the vibrations of wave on iPhone indicating Siri's speaking), those elements of nonverbal behaviors also play an vitol role in Human-Robot Interaction.

Admittedly, verbal interaction is an important element for humans communication, the majority way to seek, convey and exchange information is through speaking. Take Alexa for example, the users in the video showed some level of social interaction, mainly in verbal behavior. Users tended to use it to get general information from it, or send command to it to do something like playing a music. To be more specific, they started with greeting the robot by calling the its name following a question, like "Alexa, what's on the calendar today?". However, 4 out of 6 users interacted with Alexa without looking at Alexa, some level of regarding Alex as a tool could still be seen. In the third episode where a man asked Alex to play a jazz, he also used a gesture to indicate a social cue of "Have some music" even Alex could not see that. One man asked Alexa to ask Donimo's to send order by standing and facing towards it, showing a respect to a robot. In the fourth episode, In the last scene when the user saying goodnight to Alexa showed an additional social behavior. In terms of believability, users in the video trusted the verbal information provided without any doubts, such as in the 4th scenario when the couple debating

on the identity of the actor, they believed the answer from Alexa immediately. As for Google Home, users still had other nonverbal behaviors of Gaze, facing towards it even it can provides only graphics with 4 different color dots of nonverbal responses in this video. Since Siri has similar interface but with vibration of waves, here I only discussed Alexa and Google Home. Lastly, I was also curious how these same users will react to Alexa or Google Home when nonverbal behaviors were added to them?

However, both Alexa and Google Home provide only verbal information without taking any further embodiment into consideration. Though the behaviors of nonverbal modalities like Gaze, Gestures, Facial Expressions and so on, do not provide as rich information as verbal, apart from conveying information, they also "*regulate communication in face-to-face conversation*" mentioned in <<Embodied Conversational Agents Representation and Intelligence in User Interfaces>>. Like it said as "*when these embodied behaviors are omitted in face- to-face interaction between a user and an embodied system, users repeat themselves more and judge the system's use of language, and understanding of language, to be worse*" in <<Embodied Conversational Agents Representation and Intelligence in User Interfaces >>.

In comparison to mainly verbal responses from Alexa or Google Home, Jibo provided with more nonverbal behaviors of Gaze(also blinking, smiling eyes), Head movements. Take the receiving email scenario for example, after a message received, he raised his head towards the grandmother and she turned her head towards him without a second thought as well, showing a common social behavior of drawing and showing attentions. After the whole conversion ended, the grandmother said "Thanks Jibo" while looking at him, showing a respect for Jibo not just regarding him as a tool. This also supported the conclusion of "*Multiple nonverbal expressions of effort improve the rates of social dialogue acts for users who have already greeted the robot*" in <<Believable Robot Characters>>

Even if only concerning the verbal behaviors, Jibo's dialogue seemed more natural to human, comparing to Alexa or Google Home. For example in scenario mentioned in the last paragraph when a grandmother was cooking and received an email, Jibo notified her by starting with "Excuse me" instead of our stereotype image of robot's voice like "You have a new message" in a mundane tone, which also mentioned in <<Believable Robot Characters>> -- "*initiative in greeting users correlates with having them exhibit social attitudes toward the robot*". During this speaking, Jibo behaved more like a human by showing a pause when calculating a duration and response as "In half an hour", similar to humans who usually have a pause when some activities require more consciousness for more complex thinking. In addition, the graphics on the images vividly showed the embodiment of Jibo, like gazing, smiling eyes, animations for the girl and so forth.

Though all the robots mentioned show some extent to which they can respond personally, one important element was missing was the culturally specific character. All the scenarios in the videos were in western society, mainly in USA, where English was the major language. But, w*ow would they have a positive social impact on different societies with various cultures?* For example, Japanese have a large personal safe zone than many of other societies, they tend to bow to show greeting in formal situation, rather than just smile or wave hands; or bow to for apologies. Some society might use kissing cheeks as a way to show greeting. *How would these social behaviors in various cultures be designed in social robots?* Another example is graphs, *what graphs should be animated towards different culture societies?* To be more specific, such as the color of

Red, which Chinese regard as a color of good fortune and joy, while in modern society, red is mainly used as a sign of danger. *How would various colors affect different societies with distinct cultures?* To advancing social robots to a higher level, I believe these questions need to be addressed.

I can't predict whether Jibo would be successful in the market or not, but in the perspective of Believable Robot Characters, in summary Jibo shows more trust and social behaviors from humans than verbal-only behavior as a Robot Assistant.