Research Questions to Support Conversational Learning in the Era of Ubiquitous, Mobile Agents

Robert J. Schloss, Maria Chang, Aditya Vempaty, Arup Acharya, Ravi Kokku, Lorin Wilde, and Nirmal Mukhi
rschloss@us.ibm.com, maria.chang@ibm.com, avempat@us.ibm.com, arup@us.ibm.com,
rkokku@us.ibm.com, lorin.wilde@ibm.com, nmukhi@us.ibm.com
IBM Thomas J. Watson Research Center

Abstract: This poster explores the major question: How can conversational agents, such as smart speakers, be used for opportunistic learning that complements formal teaching? We discuss design questions that arise as devices are to be used for encouraging effective, brief conversational learning. We highlight aspects of efficacy in supporting learning through interdisciplinary study, and list research questions that need to be addressed.

Ambient conversational devices as a new interaction modality for learning

One-on-one or small group interaction of learners with skilled and sympathetic teachers has shown to increase learning outcomes. This observation served as a motivation for several research efforts in utilizing Artificial Intelligence for Education, leading to the development of sophisticated tutoring systems for one-on-one human tutoring (VanLehn, 2011). In this paper, we study the opportunities enabled by *ubiquitous conversational agents,* which can use some of the strategies employed by successful human teachers. These continuously available teaching agents, accessed at different times of the day or different places through various physical interfaces, can extend the functionality of existing intelligent tutoring systems by using multimodal signals uniquely provided by rapidly available mobile technology: location services, off-classroom time, bio-sensor data, and communication with cloud-enabled devices in the office, home, classroom, and vehicle.

The number of people of all ages using conversational agents is growing (Lenhart, 2015; Druga, Breazeal, Williams, & Resnick, 2017). Voice-based agent work builds on 50 years of technical exploration (Pieraccini, 2012). While sophisticated agents that move beyond simple information retrieval are just over the horizon, these developments are currently driven by non-learning applications. As the number of learners who access and use conversational agents outside the formal education facility grows past a tipping point, learning sciences can guide interdisciplinary research agendas and enable learning to happen *anytime* and *anywhere*.

Continuously Accessible Conversational Learning Agents (CA-CLAs)

By a learning agent, we mean software used in a learning experience, directly by a learner, and which plays a supportive role in helping the learner achieve the needed learning goals. By a Conversational Learning Agent (CLA), we mean one that can appropriately communicate to the learner in a multimodal, often speech-centric manner: it has a narrative sequence that recognizes the learner's knowledge and gaps, emotions, cultural context, and implicit or explicit goals. By a Continuously Accessible CLA (CA-CLA), we mean a CLA that can be addressed by the learner at different times and in different places, with nearly instantaneous re-start time. Some CA-CLAs retain a history of the previous conversations with the learner, while others treat each conversation as an isolated exchange of information. But all CA-CLAs can be used when a teacher is not present and, in some cases of student-initiated learning, when no teacher is even identified.

We specify research challenges that we believe are important for deploying and understanding the impact of CA-CLAs. Our research questions are driven by our own experiences building learning technology aids. We aim to stimulate discussion within the Learning Sciences community and to encourage investigations into the validity and efficacy of newly developed or proposed CA-CLAs. We believe this community is best situated to provide technologists with the design recommendations to build enjoyable and effective CLAs.

Design questions for CA-CLAs

The most basic type of CA-CLA is a conversational agent for Information Retrieval (IR). However, simply reading and hearing search results is not synonymous with learning. This raises the following ***questions***:

How can the presence of IR-enabled conversational systems promote useful learning? Should there be any limitation on how much shallow question answering can be used by the learner? Is unlimited access to agent-supplied information demotivating for mastery? If access should be limited, how should it be limited?

How can conflicting sources of information that complicate the learning process be presented? Should full and complete retrievals, including different disputed observations, interpretations and theories, be displayed to all users? How does learner expertise influence IR presentation?

More advanced CA-CLAs enable experiences beyond IR, and their effectiveness may be domain-dependent. Some subjects such as music may naturally work better as audio, while others, such as science and engineering, may naturally work better with a mix of text, visual representations, or even sketching. Several works argued for the importance of multiple representations and spatial reasoning (Wai, Lubinski, Benbow & Steiger, 2010), and auditory processing (Overy, 2000; Tallal & Gaab, 2006) across different domains. Such CA-CLA systems that employ multiple representations in multiple modalities will be required to foster sustained engagement and long-term effectiveness. ***How do we interpret the multimodal expressiveness or engagement value of a CA-CLA during learning?***

Haptic feedback has been used for simulating interactions among multiple participants at a distance and for communicating directions to individuals through wearable devices (Prasad, Taele, Olubeko, & Hammond, 2014). These technologies could be applied to CA-CLAs: to guide students through tasks that have them move, observe, or interact with their physical environment. ***What is the value of such feedback for teaching physical activities, such as sports, versus activities where the learner remains mostly stationary but uses haptic feedback to understand some physical phenomenon?***

Past works have shown the importance of affect on learning (Woolf, Burleson, Arroyo, Dragon, Cooper, & Picard, 2009) especially during prolonged learning sessions; it is less understood how affect could play a role in opportunistic learning during a multitude of short-term interactions. ***What is the most effective way to provide subtle affective feedback during learning via CA-CLAs?***

As learners hear or read new information, humans instructors can observe facial expressions or body language and interpret these non-verbal cues ("I'm getting lost, go slower"; "This is obvious, go faster") to modify information delivery. Modern methods in naturalistic environments (McDuff, El Kaliouby & Picard, 2016) suggest that scalable facial expression interpretation in mobile learning is within reach. However, when constructing mobile learning environments, we need to consider the following: ***What kind of hardware is needed to observe learner facial expression which is a feedback for CA-CLAs? How can we improve the effectiveness of interpreting and responding to non-verbal cues?***

In researching and building Conversational Learning systems at IBM, as we apply AI methodologies to the design of intelligent tutoring systems that personalize the teaching and learning experiences of disparate users under different constraints, we are gathering data and generating new questions from these real-world interactions. We will demo commercial systems for Conversational Learning during the interactive session.

References

Carnavale, A. P., Smith, N., Melton, M., & Price, E.W. (2015). Learning While Earning: The New Normal. Center on Education and the Workforce, Georgetown University.

Druga, S., Breazeal, C., Williams, R., Resnick, M. (2017, June). "Hey Google is it OK if I eat you?" Initial Explorations in Child-Agent Interaction, IDC '17.

Lenhart, A. (2015, April). Teen, Social Media and Technology Overview 2015, Pew Research Center.

McDuff, D., El Kaliouby, R., & Picard, R. W. (2015, September). Crowdsourcing facial responses to online videos. In Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on (pp. 512-518). IEEE.

Overy, K. (2000). Dyslexia, temporal processing and music: The potential of music as an early learning aid for dyslexic children. Psychology of music, 28(2), 218-229.

Pieraccini, R. (2012) The Voice in the Machine: Building Computers That Understand Speech. MIT Press.

Prasad, M., Taele, P., Olubeko, A., & Hammond, T. (2014, February). HaptiGo: A navigational 'tap on the shoulder'. In Haptics Symposium (HAPTICS), 2014 IEEE (pp. 339-345). IEEE.

Tallal, P., & Gaab, N. (2006). Dynamic auditory processing, musical experience and language development. Trends in neurosciences, 29(7), 382-390.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educational Psychologist, 46(4), 197-221.

Wai, J., Lubinski, D., Benbow, C. P., & Steiger, J. H. (2010). Accomplishment in science, technology, engineering, and mathematics (STEM) and its relation to STEM educational dose: A 25-year longitudinal study. Journal of Educational Psychology, 102(4), 860.

Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect-aware tutors: recognising and responding to student affect. International Journal of Learning Technology, 4(3-4), 129-164.

