Army researchers create pioneering approach to real-time conversational AI

April 19 2021

Army researchers create a novel approach that allows autonomous systems to flexibly interpret and respond to Soldier intent. Credit: (1st Lt. Angelo Mejia)

Spoken dialogue is the most natural way for people to interact with complex autonomous agents such as robots. Future Army operational

environments will require technology that allows artificial intelligent agents to understand and carry out commands and interact with them as teammates.

Researchers from the U.S. Army Combat Capabilities Development Command, known as DEVCOM, Army Research Laboratory and the University of Southern California's Institute for Creative Technologies, a Department of Defense-sponsored University Affiliated Research Center, created an approach to flexibly interpret and respond to Soldier intent derived from spoken dialogue with autonomous systems.

This technology is currently the primary component for dialogue processing for the lab's Joint Understanding and Dialogue Interface, or JUDI, system, a prototype that enables bi-directional conversational interactions between Soldiers and autonomous systems.

"We employed a statistical classification technique for enabling conversational AI using state-of-the-art natural language understanding and dialogue management technologies," said Army researcher Dr. Felix Gervits. "The statistical language classifier enables autonomous systems to interpret the intent of a Soldier by recognizing the purpose of the communication and performing actions to realize the underlying intent."

For example, he said, if a robot receives a command to "turn 45 degrees and send a picture," it could interpret the instruction and carry out the task.

To achieve this, the researchers trained their classifier on a labeled data set of human-robot dialogue generated during a collaborative search-and-rescue task. The classifier learned a mapping of verbal commands to responses and actions, allowing it to apply this knowledge to new commands and respond appropriately.

Researchers developed algorithms to incorporate the classifier into a dialogue management system that included techniques for determining when to ask for help given incomplete information, Gervits said.

In terms of Army impact, the researchers said this technology can be applied to combat vehicles and autonomous systems to enable advanced real-time conversational capability for Soldier-agent teaming.

"By creating a natural speech interface to these complex autonomous systems, researchers can support hands-free operation to improve situational awareness and give our Soldiers the decisive edge," Gervits said.

According to Gervits, this research is significant and unique in that it enables back-and-forth dialogue between Soldiers and autonomous systems.

"Interacting with such conversational agents requires limited to no training for Soldiers since speech is a natural and intuitive interface for humans and there is no requirement to change what they could say," Gervits said. "A key benefit is that the system also excels at handling noisy speech, which includes pauses, fillers and disfluencies—all features that one would expect in a normal conversation with humans."

Since the classifier is trained ahead of time, the system can operate in real-time with no processing delay in the conversation, he said.

"This supports increased naturalness and flexibility in Soldier-agent dialogue, and can improve the effectiveness of these kinds of mixed-agent teams," Gervits said.

Compared to commercial deep-learning approaches, which require large, expensive data sets to train the system, this approach requires orders of

magnitude fewer training examples, he said. It also has the advantage of being able to reduce deployment time and cold start capability for new environments.

Another difference is that commercial dialogue systems are typically trained in non-military domains, while his focus is on a search-and-rescue task specifically designed to mimic the style of Soldier-robot interaction that could occur in a future tactical environment.

Finally, the classification approach allows for better transparency and explainability of the system performance making it possible to analyze why the system produced a certain behavior. This is critical for military applications wherein ethical concerns demand greater transparency of autonomous systems, Gervits said.

The research was performed primarily a few years ago when Gervits was an intern at ICT. The subsequent manuscript was accepted to the International Workshop on Spoken Dialogue Systems in 2019 and presented at the conference. It was published in the conference proceedings in 2021.

Dr. David Traum, from the Natural Language Dialogue group at ICT, led the dialogue research, which included the statistical classifier. Dr. Matthew Marge from ARL led the Botlanguage project, a collaborative effort between ARL at the Adelphi Laboratory Center, ARL West and ICT.

The next steps for this research are threefold:

1. To improve the system performance by supplementing the classifier with additional linguistic representations.
2. Extending the approach to enable learning of new training

examples through real-time dialogue. An example of this is a robot encountering something new in the environment and asking a Soldier what it is.

3. Integrating additional interaction modalities such as gaze and gesture, in addition to speech, for more robust interaction in physical environments.

"With the tactical environment of the future likely to involve mixed Soldier-agent teams, I am optimistic that this technology will have a transformative effect on the future of the Army," Gervits said. "It is highly rewarding for me as a researcher to see such a tangible outcome for my efforts."

Provided by The Army Research Laboratory