The Thirty-Second AAAI Conference
on Artificial Intelligence (AAAI-18)

A Cognitive Assistant for Visualizing and Analyzing Exoplanets

Jeffrey O. Kephart, Victor C. Dibia, Jason Ellis
Biplav Srivastava, Kartik Talamadupula, Mishal Dholakia
IBM T. J. Watson Research Center, Yorktown Heights, NY 10598 USA
{kephart, dibiavc, jasone, biplavs, krtalamad, mdholak} @ us.ibm.com

Abstract

We demonstrate an embodied cognitive agent that helps scientists visualize and analyze exo-planets and their host stars. The prototype is situated in a room equipped with a large display, microphones, cameras, speakers, and pointing devices. Users communicate with the agent via speech, gestures, and combinations thereof, and it responds by displaying content and generating synthesized speech. Extensive use of context facilitates natural interaction with the agent.

Introduction

The last few years have witnessed a proliferation of voice-activated AI assistants and chatbots that help people find information, trouble-shoot problems, and perform useful tasks such as placing orders online. Inspired by the prescient vision of Licklider on human-machine symbiosis (Licklider 1960) and more recent work on symbiotic cognitive computing (Farrell et al. 2016; Kephart and Lenchner 2015), we seek to take a step beyond this current generation by creating *embodied cognitive agents*, which we define as software agents that co-inhabit a physical space with people, and use their understanding of what is happening in that space to act as valuable collaborators on cognitive tasks.

The prototype described here is an embodied cognitive assistant that helps astrophysicists visualize and analyze exoplanets — planets that orbit stars other than the sun. Since their first discovery in 1992, data on over 3600 confirmed exoplanets have been collected by ground-based and orbiting telescopes and other devices (Akeson and others 2013).

Today, astrophysicists and other scientists expend significant mental energy on the mechanics of combining, visualizing and analyzing data, and extracting knowledge from the written literature. Our goal is to create assistants that substantially reduce the cognitive burdens faced by astrophysicists and other scientists (and ultimately business people as well), allowing them to focus a greater fraction of their mental time and energy on more creative and exploratory pursuits. Our prototype takes significant steps towards this goal.

Copyright © 2018, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

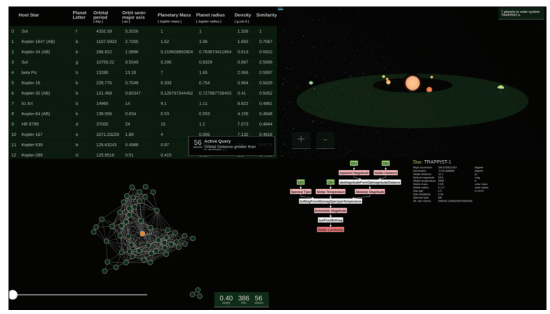

Figure 1: Snapshot of display after interactions with agent.

Technical Details

The prototype is situated in a room equipped with a large display, microphones, cameras, speakers, and pointing devices. Users communicate with the assistant via speech, gestures, and combinations thereof. The assistant responds by displaying content in the form of web pages on the screen (Fig. 1), often accompanied by synthesized speech. It can also exhibit proactive behavior, such as guiding users through the process of clarifying their requests.

The agent's perceptual system comprises two main components: speech-understanding and pointing-interpretation. To understand speech, the audio signal is fed to the IBM Watson speech engine. A conversation service receives the resulting transcript via pub-sub, sends it to the Watson Conversation service and the Watson NLU service, and combines the results to generate a data structure that contains user intent inferred from a classifier (trained to distinguish among 30 different classes), as well as entities such as physical properties, mathematical quantities, and peoples' names.

The pointing device is an HTC VIVE controller that is modified to produce a cursor that is projected on the display. The cursor's location is emitted as a mouse event, and associated with the window identity and event type. Application and spatial context are maintained in a Redis database, which supports reads and writes on a millisecond timescale.

When it receives the data structure generated by the conversation agent, the parser looks up appropriate information in the spatial and application context, and combines the received information with the context to produce a command, i.e. an environment-free JSON representation of the desired action. During this step, the originally-extracted intent may be overridden, names of properties and people may be revised, and various critical parameters may be substituted or combined. The command is sent to an application orchestrator via pub-sub. This separation of concerns enables the prototype to run in four different environments, each with different hardware devices and configurations.

When the application orchestrator receives a command, or an event signifying that a long-running computation has completed, it marshals other services and resources to carry out that command or respond appropriately to that event, culminating in displaying content on the screen as a web page and/or playing synthesized speech through the room's speakers. More detailed accounts of the orchestrator may be found in earlier papers on an embodied cognitive assistant for mergers and acquisitions (Farrell et al. 2016; Kephart and Lenchner 2015). Key advances in functionality and naturalness of interaction that distinguish the exoplanets prototype from its M&A predecessor include:

- **Deixis via simultaneous speech and gesture**. Since many astronomical objects have technical names, e.g. PSR1257+12, visualization or analysis requests that rely solely upon speech are awkward and error-prone. Our prototype supports deixis, allowing users to simultaneously point at an object and state what operation they would like to perform. For example, "tell me more about this planet", or "compute luminosity" while pointing at any representation of that object will be interpreted correctly. Some display elements (e.g. plotted points) represent both an exoplanet and its host star, in which case the agent resolves the ambiguity by consulting a list of planet and stellar properties; e.g. only stars possess "luminosity".
- **Progressive querying**. Data exploration is typically an iterative process. The user starts with a simple query, views the results, and modifies the query accordingly. When viewed in full, the resulting query can be rather complex, whereas in the mind of the user it is represented as "whatever I said before" plus a small modification. In order to match the user's mental model of the query, and to obviate excessive and awkward verbalization, a full representation is maintained in the application context, and suitably updated when a modification is received. Multiple queries can be maintained in the context simultaneously.
- **Minimizing reliance on attention word**. Using an attention word such as an avatar's name is often used to reduce the likelihood of confusing commands with side conversations with other people in the environment. This requirement proves to be annoying, especially when issuing short commands. We use context to infer when the evidence for an utterance to represent a system command is already sufficient. For example, if the user points to the x axis of a plot and says "use a log scale", the evidence is considered to be strong enough to obviate the attention word.
- **Asking for clarification**. Users do not always know the essential parameters for a given command. Accordingly, if the system receives an intent with some missing parameters, it prompts the user to supply them.
- **Explainable self-programming**. The agent uses an AI planner (Srivastava, Bigus, and Schlosnagle 2004) to program itself to derive certain physical quantities. Metadata on microfunction inputs and outputs are compiled into a PDDL domain description. When a computation is requested for a given object, its data are read from the database, and a plan manager assesses which data fields exist for that object, resulting in an input signature. The signature and the desired output constitute the planning problem statement. The planner generates a feasible plan for calculating the desired quantity, which is sent to a plan executor. The quantity resulting from the executed plan is presented verbally to the user. The plan is also stored in the application context. If the user asks "How did you compute luminosity?", the agent converts this plan into human-friendly forms that are displayed visually and played over the speaker as synthesized speech.

Functions can be invoked in any order. To generate Fig. 1, a user asked for exoplanets with orbital distance greater than Mercury, using progressive querying to narrow the search to those with mass and radius greater than Mercury (green table). Next, the user requested exoplanets most similar to Jupiter (force-directed similarity view displayed below table). Asking "Show me this planetary system" and "Tell me more about this star" while pointing at TRAPPIST-1 in the table resulted in the planetary system simulation and the star data on the right side. Finally, the user asked "Calculate luminosity" while pointing at one of the table rows, and then "How did you calculate luminosity?" — resulting in a verbal explanation plus the box diagram. For a video of the prototype, see ibm.biz/tyson-demo.

Conclusion

This prototype is a step towards our vision of embodied cognitive agents that collaborate effectively with scientists and business people. Its architecture allows it to run in four different environments with minimal customization. Additional capabilities required to make the agent a truly effective assistant include finding relevant literature pertaining to a given object or topic and extracting relevant information from the text and figures; performing regressions and other curve-fitting operations; and providing affordances that help novice users understand how to invoke its capabilities.

References

Akeson, R. L., et al. 2013. The NASA Exoplanet Archive: Data and Tools for Exoplanet Research. *PASP* 125:989.

Farrell, R.; Lenchner, J.; Kephart, J.; et al. 2016. Symbiotic cognitive computing. *AI Magazine* 37(3):81–93.

Kephart, J. O., and Lenchner, J. 2015. A symbiotic cognitive computing perspective on autonomic computing. In *ICAC 2015)*, 109–114. IEEE.

Licklider, J. C. R. 1960. Man-computer symbiosis. *IRE transactions on human factors in electronics* 1:4–11.

Srivastava, B.; Bigus, J. P.; and Schlosnagle, D. A. 2004. Bringing planning to autonomic applications with ABLE. In *Proc. ICAC*, 154–161. IEEE.