David Cohen

Carnegie Mellon University
23, NASA Research Park
Moffett Field, CA 94035

david.cohen@sv.cmu.edu

1 Research Interests

My research interests are in building intelligent computer programs which can relate to people through dialog. I think the most interesting problems related to this goal are in **dialog system architecture**, **evaluation**, and **machine teaching**. I have interest in the application of these programs toward the betterment of humankind, for example through **tutoring systems** and **personal assistants**.

1.1 Dialog System Architecture

1) What type of expertise does it take to build a dialog system? 2) What quality of dialog can you achieve per unit of training data and developer effort? 3) What set of capabilities can you implement per unit of training data and developer effort? These questions are answered if you know the answer to a fourth question: What dialog system architecture are you using?

The answers today are too often: 1) Developers need domain, linguistics, speech recognition, user experience, machine learning, and programming expertise to build a good dialog system. 2) A developer can obtain passable quality for a simple task after acquiring dozens of annotated in-domain dialogs, high task success rate and user satisfaction after collecting big data through a deployment. 3) For every question a system can answer, the developer must hard-code database queries and NLG templates. Clarification and handling of out-of-domain and out-of-capability requests must be done from scratch. Instead of building broad-capability dialog systems, engineers (and marketing teams) end up focusing on getting users to try only what is implemented. The answers should be: 1) Developers require domain expertise. 2) The architecture yields a high-quality dialog system immediately after the domain is encoded. 3) Developers can build a system capable of fluent reasoning and conversation immediately. My previous work on the YODA dialog system architecture [1] aimed to create a dialog system development process closer to this ideal.

YODA comes with generic SLU, DST, DM, and NLG components which support question-answering and command-and-control across many domains. To build a YODA dialog system requires a developer to specify the domain ontology and lexicon, implement non-dialog actions (for command-and-control), and implement the sensor interface. After this is done, they will have a fully-functioning dialog system. Sample dialog systems include an intent-launching personal assistant app reminiscent of Siri, a smart-home dialog system, and a question-answering system about Carnegie Mellon's course schedule. Future work will involve allowing developers to simply build task-oriented dialogs where the system takes the main initiative, and will support automatic language modeling.

1.2 Evaluation

Most spoken dialog system evaluation has focused on user satisfaction and task completion. The most influential such approach is the PARADISE framework (Walker et al., 1997). PARADISE presumes that maximizing user satisfaction is the purpose of dialog systems, and goes about defining an approach for predicting user satisfaction based on more readily-available performance and cost metrics.

Intelligent computer programs today, including dialog systems, tend to be very narrow in their capabilities and inflexible in their use. Part of the problem is that researchers do not have the tools to measure and report what the capabilities of their systems are, and that breadth and flexibility are not well-defined research goals. My first work to change this was the paper, (Cohen and Lane, 2016). This paper suggests a model and methodology for measuring the breadth and flexibility of a dialog system's capabilities. The approach involves having human evaluators administer a targeted oral exam to a system and provide their subjective views of that system's performance on each test problem. One goal of this work is that authors will augment their reporting with the proposed approach to improve clarity and make more direct progress toward broadly-capable dialog systems.

1.3 Machine Teaching

Machine teaching is defined in (Zhu, 2015) as the inverse problem of machine learning: finding an optimal training set to allow a learning algorithm to efficiently learn a target model. Dialog systems research has made progress toward end-to-end machine learned dialog systems where a data set is collected for the target task, such as in (Serban et al., 2016). If this work is to lead to broadly intelli-

[1] http://davidogbodfog.bitbucket.org/yoda/yoda.html

gent dialog systems, researchers will have to extend their focus beyond learning algorithms, and emphasize creating targeted training curricula to allow machine learning algorithms to learn intelligent behavior. An initial focus should be on quantifying the capabilities of machine learned dialog systems as a function of the training data used to create them.

2 Future of Spoken Dialog Research

I believe that dialog system research over the next 5-10 years will require real users and data on a scale not obtainable by academic researchers. Companies like Microsoft, Interactions, Google, and Facebook will show increased willingness to expose their research dialog systems to the public in order to make research progress. Early experiments such as Microsoft's spectacularly failed teen chatbot (Horton, 2016) exposed the shallowness of existing machine teaching and learning approaches, and will be followed up by experiments where researchers think more about the relationship between learning environment and what is learned.

Dialog state tracking will likely be standardized and generalized in the next 5-10 years, allowing task-general dialog state trackers to be trained. Standards will be developed to support more complex dialog states than current systems; for example, instead of assuming dialog state is a single object in a database, dialog state will commonly consist of more complicated data structures.

3 Suggestions for Discussion

- How should a broad-capability, open-domain dialog system be taught?
- How could we go about making tutoring systems with enough content and depth to personalize to and fully occupy a school child?

References

David Cohen and Ian Lane 2016. *An Oral Exam for Measuring a Dialog System's Capabilities.* AAAI Conference on Artificial Intelligence

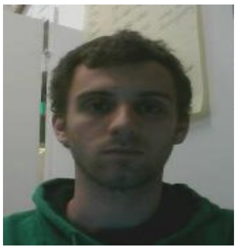
Helena Horton 2016. *Microsoft deletes 'teen girl' AI after it became a Hitler-loving sex robot within 24 hours.* The Telegraph

Serban, Iulian V., et al. 2016. *Building end-to-end dialogue systems using generative hierarchical neural network models.* AAAI Conference on Artificial Intelligence

Walker, M. A.; Litman, D. J.; Kamm, C. A.; and Abella, A. 1997. *Paradise: A framework for evaluating spoken dialogue agents.* Proc. EACL

Zhu, Jerry 2015. *Machine Teaching: An Inverse Problem to Machine Learning and an Approach Toward Optimal Education.* AAAI Conference on Artificial Intelligence

Biographical Sketch



David has received his masters degree from CMU in ECE and his bachelors from UCLA in EE. He was formerly co-founder of Podium, a dialog systems company which specialized in food ordering. He plays basketball and guitar, reads sci-fi and non-fiction, and is an outstanding dancer.