“If I Only Had a Brain”: Automation of Dialogue Management in a Wizard-of-Oz Experiment Setup

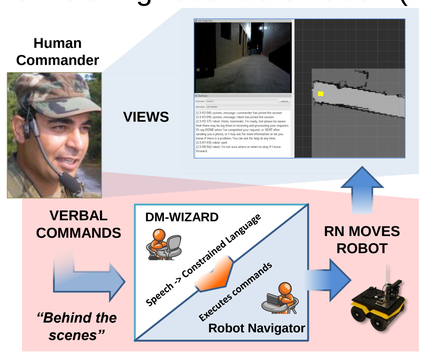
Cassidy Henry (ARL West, UCLA) – cassidy.henry@ucla.edu

Mentors: Dr. Kim Pollard (ARL West), Dr. Matthew Marge (ARL East). Collaboration on GUI work with Dr. Claire Bonial (ARL East)

GOAL: Automate wizard dialogue in a GUI to create reliable, tractable training data without sacrificing naturalness

Introduction

- Problem: We don't know how people will communicate with robots given no linguistic constraints and what the expectations are for communication in return
- We want to leverage dialogue management approaches from human-virtual human dialogue and apply it to HRI
- We need to balance naturalness with functionality (e.g. leverage human conversational techniques such as those described in *Jurafsky & Martin 2007)*
- Experiment is a two-person Wizard of Oz setup; it is beneficial to have two people performing the task of simulating robot automation (*Marge et al. 2016a)*



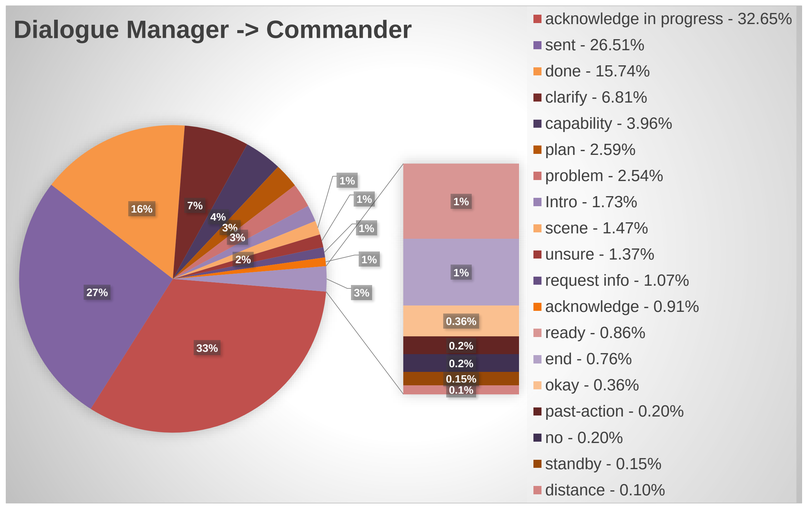
The Experiment:
Collaborative search and navigation task with a robot teammate, where robot is directed using natural language as opposed to teleoperation

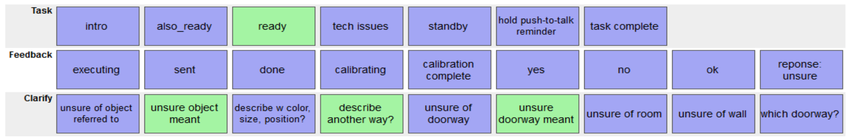
10 participants engaged in three separate 20 minute tasks with a robot (Dialogue Manager Wizard and Robot Navigator, participant never interfaces with RN directly), leading to 30 experiment runs with 2 bidirectional channels of communication

Analysis

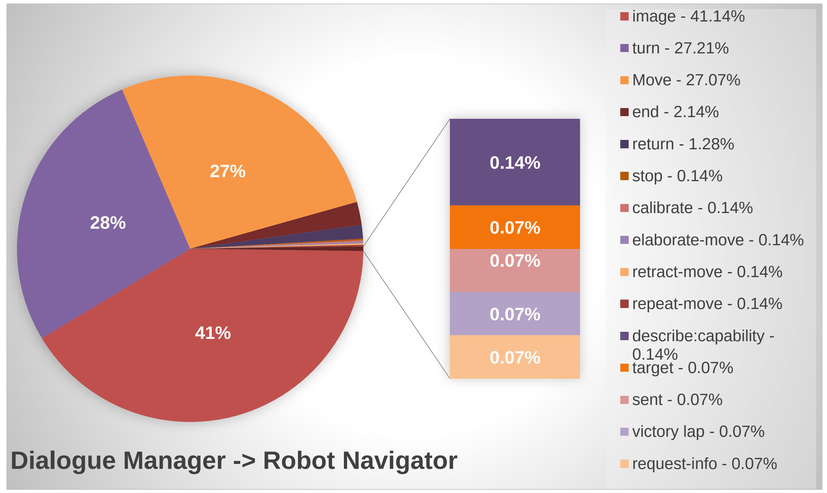
- All 60 exchanges were first annotated according to dialogue moves (*Marge et al. 2016b)*
- Validated and expanded set of dialogue moves as needed
- Frequency analyses were then performed on dialogue move parameters, dialogue move types, and unique utterance occurrences
- In dialogue parameters, only few occurred frequently: left/right, forward, 45 and 90 degrees, etc) – high variability in speech data!
- Non-singleton utterances have full coverage in the current version of the interface (53 unique utterances covers *85.85% of total* – 1692/1971 total utterances)

Data





Pictured are the analyses performed on counts of communications of the dialogue manager. The counts are of annotations of dialogue moves. These frequency analyses combined with the “wizardsourced” data served as basis for developing the GUI.



Discussion

- Our experiment methodology differs from Q-A type systems in that it strives to be conversationally interactive with humans and contextually aware of its environment at large
- GUI was created to address our experiment's goal towards movement to automation
- Delicate balance must be considered between technical constraints and conversational techniques
- Synonymous wordings also present their own unique issue - *context dependent*
- To solve this, we use careful consideration when deciding if coverage for an utterance has been achieved in the GUI

Steps Forward

- Stress testing the GUI using aligned language data
- Analysis of coverage for singleton utterances
- Assess impacts of automated vs human dialogue manager after Experiment 2 (which is intended to assess any impact on dialogue exchanges from the graphical interface vs free response typing).

References

Dan Jurafsky and James H Martin. 2007. *Speech and language processing*. Pearson Prentice Hall, Upper Saddle River, N.J., edition.

Matthew Marge, Claire Bonial, Brendan Byrne, Taylor Cassidy, A. William Evans, Susan G. Hill and Clare Voss. 2016. Applying the Wizard-of-Oz Technique to Multimodal Human-Robot Dialogue. In Proc. of IEEE RO-MAN, Columbia University, Teachers College.

Matthew Marge, Claire Bonial, Kimberly A. Pollard, Ron Artstein, Brendan Byrne, Susan G. Hill, Clare Voss, and David Traum. 2016. Assessing Agreement in Human-Robot Dialogue Strategies: A Tale of Two Wizards. In Proc. of IVA.

Acknowledgements

First and foremost, I would like to thank Dr. Claire Bonial for her hard work in taking my analyses and transforming them into the GUI. Additional thanks go out to Dr. Kim Pollard and Dr. Matthew Marge for being awesome and incredibly helpful mentors to me at ARL, and to Dr. David Traum and Dr. Ron Artstein at ICT for their input and feedback during my work and analyses of the communications analyzed for this research.

This work was completed as a visiting research assistant funded by the U.S. Army Research Lab; anything expressed in this material is the opinion of the author and does not necessarily reflect the views of the U.S. Government and no official endorsement should be inferred.

APPROVED FOR PUBLIC RELEASE – DISTRIBUTION UNLIMITED