Are We Conversational Yet?

A Design Study And Empirical Evaluation of Multi-Turn Dialogues For Virtual Assistants

Project Pitch – CS294S Fall 2020
Almond is out there, now what?

- Almond 1.99 released in September 2020
- First assistant to support multi-turn dialogues using a *contextual neural network*
- Automatically generated replies, suggestions and follow-ups
- So we’re done right?

- Spoiler: *Almond doesn’t work*
Happy vs. Unhappy Paths

● Wizard-of-Oz dialogues are mostly happy paths
  ○ Both the agent and user have a common goal of completing the transactions
  ○ They are playing along with no surprises and with no “computer errors”

● 90-10 rule in software engineering:
  ○ We need to spend 90% of the effort to handle the last 10% (due to exception handling)
  ○ In NLP dialogues, given the expected failures in NLP, this is higher.

● What are possible causes of unhappy paths?
Modularizing The State Machine

- Developers concentrate on the application-specific logic
- Common modules take care of completing a “command”
  - E.g. Slot filling is a “mini-dialogue”
    - inserted for every incomplete request

- Model the major unhappy reasons and alternative paths abstractly
Challenges

● How do we control the dialogue agent to minimize unexpected answers?
  ○ User studies to evaluate different kinds of agent responses.

● What methodology can we use to identify the abstract dialogue acts in unhappy paths?
  ○ Are there transcripts? How do human agent transcripts compare with AI agent transcripts.
  ○ Can we role play? Can we crowdsourcing at scale?
  ○ Can we assume that language variations with the same intent can be handled automatically? (like auto-QA)
  ○ Hypothesis: the first 70% is easy; the rest needs iterative refinement after deployment. Tools are necessary.

● Can we create a “backoff” scheme, such as reading the possible choices that the agent can understand? (like a menu)
High-level Project Plan

- **Step 0:** Familiarize with existing Almond

- **Step 1:** Pilot study to identify happy and unhappy paths
  - Small scale crowdworker test or even with friends and family

- **Step 2:** Expand (or contract) dialogue capabilities to improve success ratio

- **Step N:** Iterate until success

- **Step N+1:** Profit!
Schedule

- Create a strawman of possible abstract states (2 weeks)
  - Test Almond to get an intuitive feel
  - Try a small-scale formative study to gauge user responses.

- Design a crowdsourcing experiment for a small domain (2 weeks)

- If the results are reasonable,
  - implement a subset of the dialogue and test on users (2 weeks);
  - If not, try another experiment.
Why You Should Work on This Project

- **Dialogues are the next big thing for assistants**
  - We all experience really bad customer support over the phone!
  - The first round is the low hanging fruit.

- **We have a secret weapon**
  - The contextual neural network is our state of the art model nobody else has.

- **Get To Research Quick**: infrastructure is already built