Are We Conversational Yet?
A Design Study And Empirical Evaluation of Multi-Turn Dialogues For Virtual Assistants

CS 294S/W Research Project Proposal
Alejandrina G.R., Kat McNeill
What will virtual assistants look like in ~5 years?
What are current virtual assistants lacking?
Are virtual assistants actually conversational?
Not conversational yet.

And we don’t really know if people want dialogues to begin with.
Motivation

User-centered 🧑💻

Infrastructure already built 🙌

Multi-turn dialogues with contextual neural network 👀

Dialogues = next big thing for virtual assistants? 😞

Bottleneck for Almond 😞
Definitions

**Dialogue:** well-formed sequence of turns
**State:** formal representation of the dialogue, up to a certain point
**User state:** right after the user speaks
**Result state:** right after execution
**Agent state:** right after the agent speaks
**Abstract State:** family of states, as defined by the dialogue state machine
Current state

30 state transactional models.
All transitions are now built with artificial dataset.
Built in a uniform way.
Big claim: 30 transactions cover all the domains.

State-Machine-Based Dialogue Agents
with Few-Shot Contextual Semantic Parsers

Giovanni Campagna  Sina J. Semnani  Ryan Kearns  Lucas Jun Koba Sato  Monica S. Lam
Computer Science Department
Stanford University
Stanford, CA, USA
{gcampagn,sinaj,kearns,satoj,klam}@cs.stanford.edu

Abstract
This paper presents a methodology and toolkit for creating a rule-based multi-domain conversational agent for transactions from (1) language annotations of the domains’ database schemas and APIs, and (2) as trees of conversations. The leaves of the tree represents an agent utterance and a small set of intents capturing the anticipated user responses (Gao et al., 2018). Wizard-of-Oz conversations (Kelley, 1984) have traditionally been used to study dialogue state tracking. We realize that they can be put to use to create
State-Machine-Based Dialogue Agents with Few-Shot Contextual Semantic Parsers

Giovanni Campagna  Sina J. Semnani  Ryan Kearns  Lucas Jun Koba Sato  Monica S. Lam
Computer Science Department
Stanford University
Stanford, CA, USA
{gcampagn,sinaj,kearns,satoj,k,lm}@cs.stanford.edu

Abstract
This paper presents a methodology and toolkit for creating a rule-based multi-domain conversational agent for transactions, from (1) a large language model trained on large amounts of dialogue data, (2) a few-shot contextual parser to identify high-level intents in user utterances, and (3) a grammar that specifies the structure of the tree representing an agent utterance and a small set of intents capturing the anticipated user responses (Gao et al., 2018).
Unhappy paths:
How to deal with things we have never seen before?
Unhappy Paths

How will the user respond to unexpected answers?

Can we anticipate how people might do multi turn dialogues?

The first sentence that is not understood the agent is already off track.

Something we have never modeled.
How do we minimize unexpected answers? Can we crowdsource at scale?

What methodology is best to identify the abstract dialogue acts in unhappy paths?

How do human agent transcripts compare with AI agent transcripts? Can we use a “backoff” scheme?

Can we assume that language variations with the same intent can be handled automatically?

What are the different things that people want to say?
01. Test Almond to get an intuitive feel
02. Run a small-scale formative study to gauge user responses. (Use GTP-2 Dialogue?)
03. Design a crowdsourcing experiment for a small domain

Timeline

04. If good results from 3, improve success ratio. Else try another experiment.
05. Expand dialogue capabilities and iterate.
06. Final demo/presentation and writeup.
Demo

Results from studies

Expanded dialogue capabilities

Aiming for 50 dialogue transitions (now we have 30)
Sources

- A Formal Language & Data Engineering Approach to Multi-Turn Dialogues
- Genie: A Generator of Natural Language Semantic Parsers for Virtual Assistant Commands
Questions?