Understanding User’s Cross-Domain Intentions in Spoken Dialog Systems

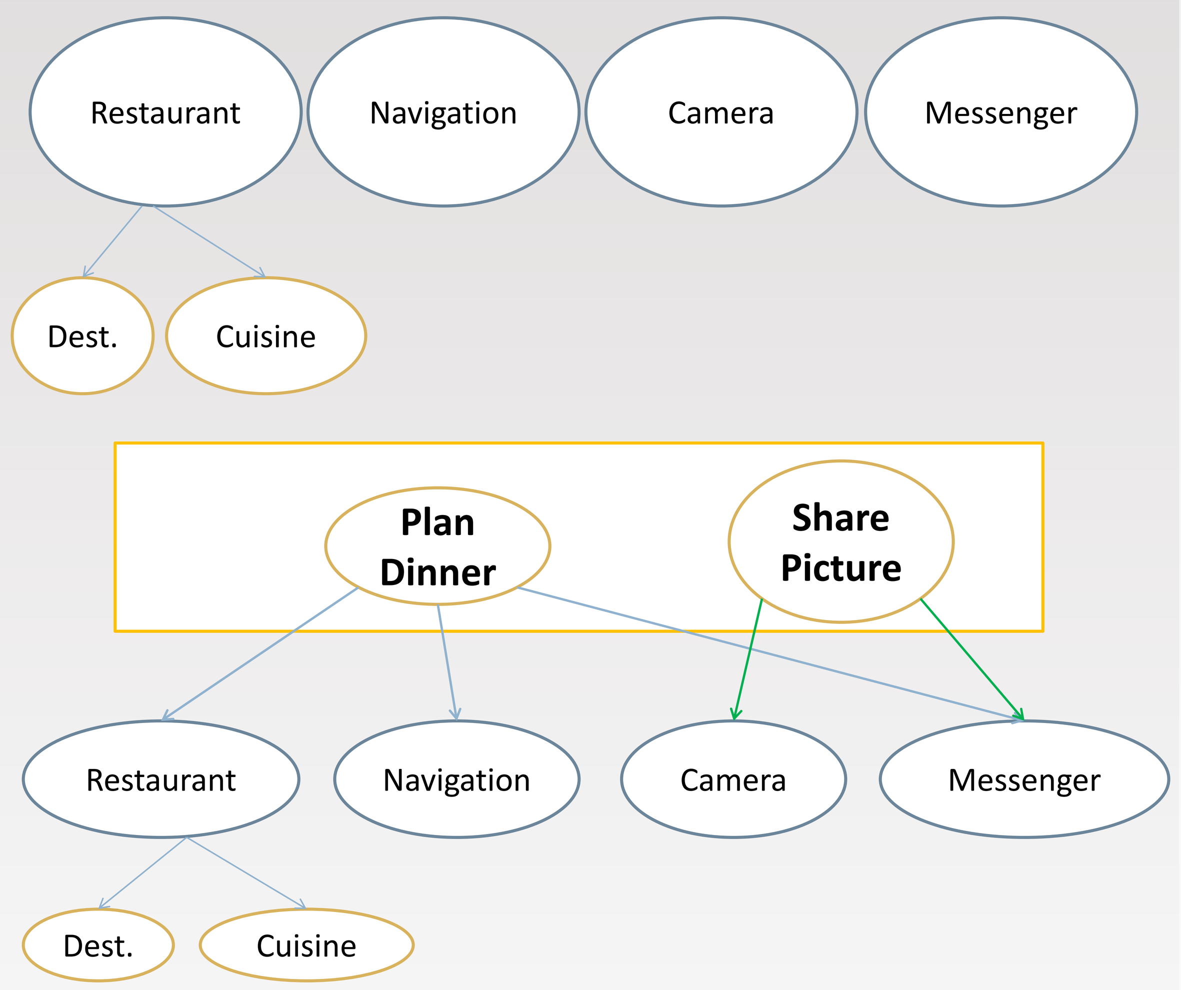
Ming Sun, Yun-Nung Chen and Alexander I. Rudnicky

School of Computer Science, Carnegie Mellon University

ABSTRACT

Spoken language based intelligent assistants (IAs) have been developed for a number of domains but their functionality has mostly been confined to the scope of a given app. One reason is that it's is difficult for IAs to infer a user's intent without access to relevant context and unless explicitly implemented, context is not available across app boundaries. We describe context-aware multi-app dialog systems that can learn to 1) identify meaningful user intents; 2) produce natural language representation for the semantics of such intents; and 3) predict user intent as they engage in multi-app tasks. As part of our work we collected data from the smartphones of 14 users engaged in real-life multi-app tasks. We found that it is reasonable to group tasks into high-level intentions. Based on the dialog content, IA can generate useful phrases to describe the intention. We also found that, with readily available contexts, IAs can effectively predict user's intents during conversation, with accuracy at 58.9%.

SINGLE-DOMAIN vs. MULTI-DOMAIN TASKS



MOTIVATION

1. Human have complex tasks that span multiple domains.
2. Human can do it.
3. Why can’t an agent?

OBJECTIVES

1. Discover the basic multi-domain user intentions.
2. Understand the current multi-domain intention and express it via natural language.

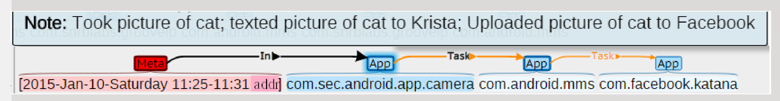
DATA COLLECTION

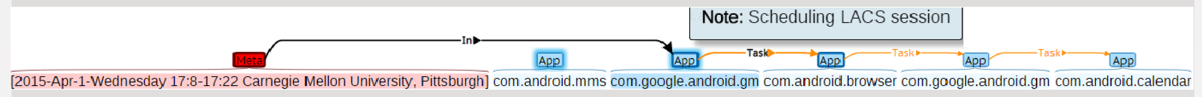
Phase-1: Collect real-life multi-domain interactions from user’s smart phone (GUI)

Phase-2: Let user talk to a Wizard-of-Oz system to reproduce the interactions via speech.

Smart Phone Log Annotation

1. Log app invocation + contexts (e.g., location, time, date)
2. Segment log into episodes if there is 3 min inactivation
3. Let user annotate each episode in terms of:
    1. Task Structure
    2. Task Description

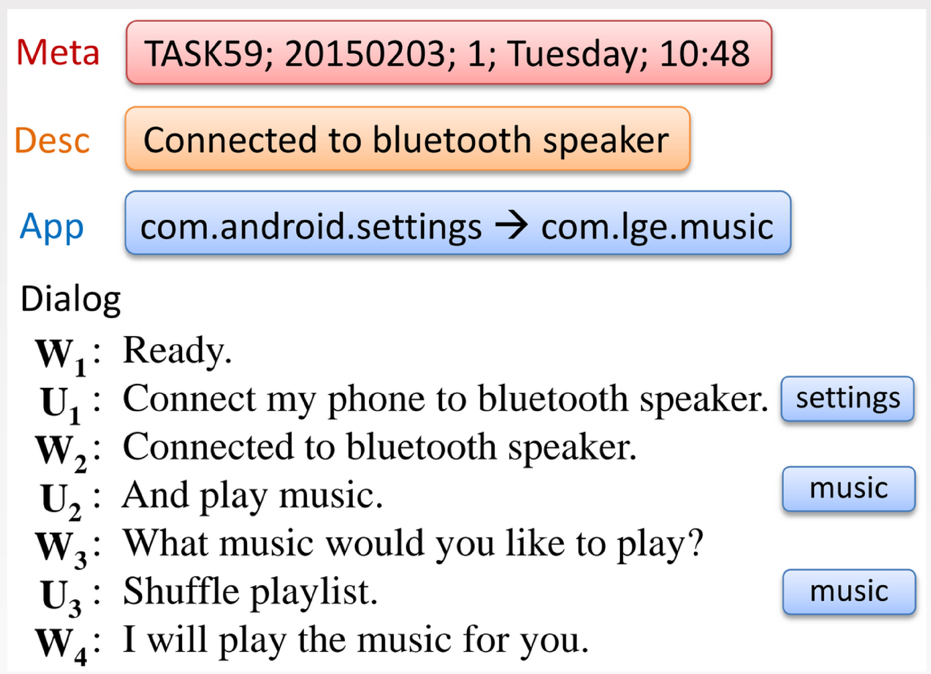




Wizard-of-Oz Dialogs

The following contexts are shown to the user and the wizard:

1. Location, time, date
2. Task description provided by the user
3. Sequence of apps used in the smart phone interaction (not required to follow the exact order of apps)



Data Statistics

- 14 participants over 1-3 months
    - avg. age 31 year-old
    - 4 male
    - 12 native speakers of English
- 533 parallel interactions in total
- 455 interactions involve more than one user turn
- 1607 user turns

EXPERIMENTS

Main Objectives:

1. **App-level:** Predict the next app via contexts
2. **Intention-level:**
    1. Discover basic intentions
    2. Refer to the intention via natural language
    3. Recognize the current intention

Exp 1: Context-aware App Prediction

Objective: Use the following contexts (or combinations) in previous turn to predict the user’s current need for an app.

- Language (user utterance, e.g., “connect my phone to the Bluetooth speaker”)
- App
- Meta: time, day, location

Method: Multi-class classification

Metrics: Top-1 Accuracy (ACC), Mean Average Precision (MAP)

Baseline: Majority (23.9%ACC, 31.7% MAP)

Result: **Using Language and App contexts together can significantly improve the performance** (40.0%ACC, 45.0%MAP).

Exp 2.1: Intention Discovery

Objective: Find basic high-level user intentions from previous interactions.

Method: KMeans clustering based on the following features:

- Task description
- User utterances
- Apps involved
- Meta: time, day, location

Metrics: subjective evaluation

Agreement with “the dialogs shown in this cluster are essentially the same task” (1-5 scale)

Result: **Users like the clustering** (4.2±1.2 out of 5.0)

Cluster	Example
1	“Picture messaging XX”, “Take picture and send to XX”
2	“Look up math problems”, “Doing physics homework”, “Listening to and trying to buy a new song”
3	“Talking with XX about the step challenge”, “Looking at my step count and then talking to XX about the step challenge”
4	“Playing [game] Spiderman”, “Allocating memory for Spiderman”
5	“Using calculus software”, “Purchasing Wolfram Alpha on the play store”
6	“Texting and calling XX”, “Ask XX if she can talk then call”

Exp 2.2: Intention Representation

Objective: Enable the agent to reveal its understanding of the recognized intention via natural language, e.g., “I think you want to *picture message somebody*.”

Method: summarize the interactions in the same cluster with key phrases.

Resources: 1. task descriptions (Desc); 2. user utterances (ASR or Manual); 3. Desc + ASR; 4. Desc + Manual.

Metrics: 1. Mean Reciprocal Rank,; 2. Mean Average Precision at K; 3. Precision at K.

Result:

- **Users find language reference understandable** (MRR=0.6).
- **Descriptions is useful resource.**

Exp 2.3: Intention Recognition

Objective: Use contexts (or combinations) in previous turn to recognize the user’s current intention.

Method: Multi-class classification

Metrics: Top-1 Accuracy (ACC), Mean Average Precision (MAP)

Baseline: Majority (31.7%ACC, 44.4% MAP)

Result: **Using Language, App and Meta contexts together can significantly improve the performance** (58.9%ACC, 66.1%MAP).

CONCLUSION

- We conducted a user study to investigate how human could interact with Intelligent Agents via speech in activities that span multiple domains/apps.
- We demonstrated that unsupervised techniques can be used for automatically identifying high-level user intentions from dialogs, task descriptions, and other sources.
- Language resources such as user utterances, task descriptions can generate understandable reference to intentions.
- Using context-aware machine learning models, the agent can predict user’s need for an app and recognize the user’s high-level intentions.

REFERENCES

- Ming Sun, Yun-Nung Chen and Alexander I. Rudnicky. HELPR: A Framework to Break the Barrier across Domains in Spoken Dialog Systems. (to appear) IWSDS-2016.
- Yun-Nung Chen, Ming Sun and Alexander I. Rudnicky. Leveraging behavioral patterns of mobile applications for personalized spoken language understanding. ICMI-2015.
- Choonsung Shin, Jin-Hyuk Hong, and Anind K. Dey. Understanding and prediction of mobile application usage for smart phones. UbiComp-2012.
- Claudio Lucchese, Salvatore Orlando, Raffaele Perego, Fabrizio Silvestri, and Gabriele Tolomei. Identifying task-based sessions in search engine query logs. WSDM-2011.
- Michael W. Berry and Jacob Kogan. Text mining: applications and theory. 2010.
- Robert Tibshirani, Guenther Walther, and Trevor Hastie. Estimating the number of clusters in a data set via the gap statistic. Royal Statistical Society: Series B (Statistical Methodology), 2001.

ACKNOWLEDGEMENT

We would like to thank Yahoo! and General Motors Advanced Technical Center for providing funding. We also thank Zhenhao Hua, Yulian Tamres-Rudnicky and Arnab Dash for their help in the data collection.