Leveraging Behavioral Patterns of Mobile Applications for Personalized Spoken Language Understanding

YUN-NUNG (VIVIAN) CHEN
MING SUN
ALEXANDER I RUDNICKY
ANATOLE GERSHMAN
Introduction

- Global Digital Statistics (2015 Jan)

<table>
<thead>
<tr>
<th>Global Population</th>
<th>Active Internet Users</th>
<th>Active Social Accounts</th>
<th>Active Mobile Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.21B</td>
<td>3.01B</td>
<td>2.08B</td>
<td>3.65B</td>
</tr>
</tbody>
</table>

Spoken language interfaces become important are incorporated in smart devices as intelligent assistants (IAs).

Spoken language understanding (SLU) is a key component of IA, which predicts users’ intended apps by understanding input utterances.
Introduction

• Task: user intent prediction
• Challenge: language ambiguity

1. User preference
   ✓ Some people prefer “Message” to “Email”
   ✓ Some people prefer “Outlook” to “Gmail”

2. App-level contexts
   ✓ “Message” is more likely to follow “Camera”
   ✓ “Email” is more likely to follow “Excel”

Considering behavioral patterns in history to model SLU for intent prediction.
Data Collection

• Subjects’ app invocation is logged on a daily basis

• Subjects annotate their app activities with
  – **Task Structure**: link applications that serve a common goal
  – **Task Description**: briefly describe the goal or intention of the task

<table>
<thead>
<tr>
<th>Meta</th>
<th>TASK59; 20150203; 1; Tuesday; 10:48</th>
</tr>
</thead>
<tbody>
<tr>
<td>App</td>
<td>com.android.settings → com.lge.music</td>
</tr>
<tr>
<td>Desc</td>
<td>play music via bluetooth speaker</td>
</tr>
</tbody>
</table>

• Subjects use a wizard system to perform the annotated task by speech

<table>
<thead>
<tr>
<th>Dialogue</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>W</strong>₁:</td>
<td>Ready.</td>
</tr>
<tr>
<td><strong>U</strong>₁:</td>
<td>Connect my phone to bluetooth speaker.</td>
</tr>
<tr>
<td><strong>W</strong>₂:</td>
<td>Connected to bluetooth speaker.</td>
</tr>
<tr>
<td><strong>U</strong>₂:</td>
<td>And play music.</td>
</tr>
<tr>
<td><strong>W</strong>₃:</td>
<td>What music would you like to play?</td>
</tr>
<tr>
<td><strong>U</strong>₃:</td>
<td>Shuffle playlist.</td>
</tr>
<tr>
<td><strong>W</strong>₄:</td>
<td>I will play the music for you.</td>
</tr>
</tbody>
</table>
SLU for Intent Prediction

User Utterance
- take this photo
- tell vivian this is me in the lab
- check my grades on website
- send an email to professor
- take a photo of this
- send it to alex

Intended App
- CAMERA IM
- CHROME EMAIL

Issue: unobserved hidden semantics may benefit understanding

Solution: use matrix factorization to complete a partially-missing matrix based on a low-rank latent semantics assumption.
Matrix Factorization (MF)

- The decomposed matrices represent low-rank latent semantics for utterances and words/histories/apps respectively.
- The product of two matrices fills the probability of hidden semantics.
Parameter Estimation

• Model implicit feedback by completing the matrix
  – not treat unobserved facts as negative samples (true or false)
  – give observed facts higher scores than unobserved facts

\[
\begin{align*}
  f^+ &= \langle u, x^+ \rangle \\
  f^- &= \langle u, x^- \rangle \\
  p(f^+) &> p(f^-)
\end{align*}
\]

\[
p(M_{u,x} = 1 \mid \theta_{u,x}) = \sigma(\theta_{u,x}) = \frac{1}{1 + \exp(-\theta_{u,x})}
\]

• Objective:

\[
\sum_{f^+ \in \mathcal{O}} \sum_{f^- \notin \mathcal{O}} \ln \sigma(\theta_{f^+} - \theta_{f^-})
\]

– the model can be achieved by SGD updates with fact pairs

The objective is to learn a set of well-ranked apps per utterance.
SLU Modeling by MF

User Utterance
- *take this photo*
- *tell vivian this is me in the lab*
- *check my grades on website*
- *send an email to professor*
- *take a photo of this*
- *send it to alex*

Intended App
- *CAMERA IM*
- *CHROME EMAIL*

Reasoning with Matrix Factorization for Implicit Intents
Experiments

- Dataset: 533 dialogues (1,607 utterances); 455 multi-turn dialogues
- Google recognized transcripts (word error rate = 25%)
- Evaluation metric: accuracy of user intent prediction (ACC) mean average precision of ranked intents (MAP)
- Baseline: Maximum Likelihood Estimation (MLE)

<table>
<thead>
<tr>
<th>Approach</th>
<th>Lexical</th>
<th>Behavioral</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) MLE</td>
<td>User-Indep</td>
<td></td>
<td>13.5 / 19.6</td>
</tr>
<tr>
<td>(b)</td>
<td>User-Dep</td>
<td></td>
<td>20.2 / 27.9</td>
</tr>
</tbody>
</table>

The user-dependent model is better than the user-independent model.
Experiments

- Dataset: 533 dialogues (1,607 utterances); 455 multi-turn dialogues
- Google recognized transcripts (word error rate = 25%)
- Evaluation metric: accuracy of user intent prediction (ACC) and mean average precision of ranked intents (MAP)
- Baseline: Maximum Likelihood Estimation (MLE) and Multinomial Logistic Regression (MLR)

<table>
<thead>
<tr>
<th>Approach</th>
<th>Lexical</th>
<th>Behavioral</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) MLE</td>
<td>User-Indep</td>
<td>13.5 / 19.6</td>
<td></td>
</tr>
<tr>
<td>(b)</td>
<td>User-Dep</td>
<td>20.2 / 27.9</td>
<td></td>
</tr>
<tr>
<td>(c) MLR</td>
<td>User-Indep</td>
<td>42.8 / 46.4</td>
<td>14.9 / 18.7</td>
</tr>
<tr>
<td>(d)</td>
<td>User-Dep</td>
<td>48.2 / 52.1</td>
<td>19.3 / 25.2</td>
</tr>
</tbody>
</table>

Lexical features are useful to predict intended apps for both user-independent and user-dependent models.
Experiments

- Dataset: 533 dialogues (1,607 utterances); 455 multi-turn dialogues
- Google recognized transcripts (word error rate = 25%)
- Evaluation metric: accuracy of user intent prediction (ACC) and mean average precision of ranked intents (MAP)
- Baseline: Maximum Likelihood Estimation (MLE) and Multinomial Logistic Regression (MLR)

Combining lexical and behavioral features improves performance of the MLR model, which models explicit information from observations.
Experiments

- Dataset: 533 dialogues (1,607 utterances); 455 multi-turn dialogues
- Google recognized transcripts (word error rate = 25%)
- Evaluation metric: accuracy of user intent prediction (ACC) mean average precision of ranked intents (MAP)
- Baseline: Maximum Likelihood Estimation (MLE) Multinomial Logistic Regression (MLR)

<table>
<thead>
<tr>
<th>Approach</th>
<th>Lexical</th>
<th>Behavioral</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) MLE (b)</td>
<td>User-Indep</td>
<td>13.5 / 19.6</td>
<td></td>
</tr>
<tr>
<td>(b) User-Dep</td>
<td>20.2 / 27.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(c) MLR (d)</td>
<td>User-Indep</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(d) User-Dep</td>
<td>42.8 / 46.4</td>
<td>14.9 / 18.7</td>
<td>46.2⁺ / 50.1⁺</td>
</tr>
<tr>
<td>(e) (c) + Personalized MF</td>
<td>47.6 / 51.1</td>
<td>16.4 / 20.3</td>
<td>50.3** / 54.2**</td>
</tr>
<tr>
<td>(f) (d) + Personalized MF</td>
<td>48.3 / 52.7</td>
<td>20.6 / 26.7</td>
<td>51.9** / 55.7**</td>
</tr>
</tbody>
</table>

Personalized MF significantly improves MLR results by considering hidden semantics.
Conclusion

• An MF model exploits both lexical and behavioral features for SLU, which considers **implicit semantics** to **enhance intent inference** given the noisy ASR inputs.

• We are able to model users’ **contextual behaviors** and their **app preference** for better intent prediction.

• The proposed multi-model personalized SLU effectively improves intent prediction performance, achieving about **52% on turn accuracy** and **56% on mean average precision** for ASR transcripts with 25% word error rate.
THANKS FOR ATTENTIONS!

Q&A

Data Available at
http://AppDialogue.com