SkillExplorer: Understanding the Behavior of Skills in Large Scale

Zhixiu Guo, Zijin Lin, Pan Li, Kai Chen

SKLOIS, Institute of Information Engineering, Chinese Academy of Sciences, China
School of Cyber Security, University of Chinese Academy of Sciences, China
Background

Virtual personal assistant (VPA)
Background

Virtual personal assistant (VPA)
Background

Virtual personal assistant (VPA)

Preset

Third-party skills
Motivation

Skills are not safe enough

Alexa, open cat fact

User

Speech recognition systems

Skill

Cat fact

Cat fax

Skill Squatting Attacks on Amazon Alexa

Motivation

• A skill is a voice app
• Recent works focus on invocation mechanism but not the content of skills

If we can analyze the content of a skill just like analyzing a traditional app.

✓ Systematic
✓ Automatic
Motivation

• A skill is a voice app
• Recent works focus on invocation mechanism but not the content of skills

If we can analyze the content of a skill just like analyzing a traditional app.

✓ Systematic
✓ Automatic

SkillExplorer
Challenges

Traditional apps and voice apps: similar but have essential differences.
Traditional Apps V.S. Voice Apps

installation

usage

analysis
Traditional Apps V.S. Voice Apps

installation

usage

analysis
Traditional Apps V.S. Voice Apps

- installation
- usage
- analysis
Traditional Apps V.S. Voice Apps

- Installation
- Usage
- Analysis

Alexa, open/⋯ + invocation names
Traditional Apps V.S. Voice Apps

- installation
- usage
- analysis

- Alexa, open/… + invocation names
- code
- document

???
Challenges

• Fully black-box

• Inputs/outputs of skills are in the form of natural languages
Challenges

• Fully black-box

• Inputs/outputs of skills are in the form of natural languages

A online chatbot Mitsuku
Approach: SkillExplorer

1. Utterances extraction
2. Interactive system
   - Question understanding
3. Answers generation
4. Skill exploration
   - Simulator
   - Utterance corpus
   - Skill
Approach: SkillExplorer

1. Utterances extraction
2. Interactive system
   - Question understanding
3. Answers generation
4. Skill exploration

Utterance corpus

Simulator

Skill
Approach: SkillExplorer

1. Utterances extraction
   - “Alexa, open Washington Post”
   - “Alexa, ask Washington Post for politics”
   - “Alexa, ask Washington Post for Post Reports”

2. Utterance corpus

3. Answers generation

4. Simulator
   - Skill
Approach: SkillExplorer

1. Utterances extraction
2. Interactive system
   - Question understanding
3. Answers generation
4. Skill exploration

Utterance corpus

Simulator

Skill
Interactive system—Five question types

- Yes/No
- Instruction
- Selection
- Wh
- Mix

Basic Corpus of Replies
Interactive system——Five question types

Q: Are you ready?
A: [yes, no]

A constituency-based parse tree
Interactive system——Five question types

Q: For any information on how to use the skill, just say: Help me.

A: [help me]

***Rules to generate answers for Instruction questions***

<table>
<thead>
<tr>
<th>ASK</th>
<th>SAY</th>
</tr>
</thead>
<tbody>
<tr>
<td>ask (sb.) Wh-Q</td>
<td>say Wh-Q</td>
</tr>
<tr>
<td>ask sth. like/... INS</td>
<td>say sth. like/... INS</td>
</tr>
<tr>
<td>ask (sb.) to INS</td>
<td>say INS (to do sth)</td>
</tr>
<tr>
<td>ask (sb.) (about/for) INS</td>
<td>say INS for sth</td>
</tr>
<tr>
<td>ask that INS</td>
<td>say (that) INS</td>
</tr>
</tbody>
</table>
Interactive system——Five question types

Selection_SC (sequence numbers or sequence letters)

A: [one, two, three]

Selection_CC (coordinating conjunctions label)

Q: To get started, you can get a quote or listen to the daily briefing.
A: [get a quote, listen to the daily briefing]
To get started, you can get a quote or listen to the daily briefing.

A: [get a quote, listen to the daily briefing]
Interactive system——Five question types

Q: What is your gender?
A: [male, female]

Virtual Users database

<table>
<thead>
<tr>
<th>Info</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Name</td>
<td>James C Washington</td>
</tr>
<tr>
<td>Gender</td>
<td>male</td>
</tr>
<tr>
<td>Date of Birth</td>
<td>6/19/1980</td>
</tr>
<tr>
<td>Phone Number</td>
<td>716-780-4085</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

An example of the virtual user
Interactive system—Five question types

Q: Please respond by saying lenses or glasses.
A: [lenses, glasses]

<table>
<thead>
<tr>
<th>Rule</th>
<th>Situation</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>$\exists Y$</td>
<td>Y</td>
</tr>
<tr>
<td>R2</td>
<td>$\exists S_{SC} &amp; \exists I$</td>
<td>S_SC&amp;I</td>
</tr>
<tr>
<td>R3</td>
<td>$(I&amp;S_{CC})$ in $Q$ $*$</td>
<td>I&amp;SC_CC</td>
</tr>
<tr>
<td>R4</td>
<td>$\exists I$</td>
<td>I</td>
</tr>
<tr>
<td>R5</td>
<td>$\exists S$</td>
<td>S</td>
</tr>
</tbody>
</table>

Rules to generate answers for Mix questions
Approach: SkillExplorer

1. Utterances extraction
2. Interactive system
   - Question understanding
3. Answers Generation
4. Utterance corpus
   - Simulator
   - Skill exploration
   - Skill
Approach: SkillExplorer

An interactive tree (i-tree for short) is used to record the exploration
• Each node represents a single interaction (include an input and a output)
• Different answers produce different branches
• The node will be marked as visited if it is explored
• Re-start from the beginning to the unvisited nodes
Approach: SkillExplorer

An interactive tree (i-tree for short) is used to record the exploration
- Each node represents a single interaction (include an input and a output)
- Different answers produce different branches
- The node will be marked as visited if it is explored
- Re-start from the beginning to the unvisited nodes

An example of i-tree

I: about us
O: …Would you like to learn more?

I: yes
O: …

I: open c. s. n. premier collision
O: … To learn more, say, about us, services, mobile app, phone number, address, or website. You can say, repeat or stop, at any time.
**Speed up mechanism**

- ignore the same questions (i.e. outputs) in different nodes
- not wait for every output reading

---

**I:** about us  
**O:** ... Would you like to learn more?

**I:** yes  
**O:** ... To learn more, say, about us, services, mobile app, phone number, address, or website. You can say, repeat or stop, at any time.
I: open c. s. n. premier collision
O: … To learn more, say, about us, services, mobile app, phone number, address, or website. You can say, repeat or stop, at any time.

I: about us
O: … Would you like to learn more?

I: yes
O: … To learn more, say, about us, services, mobile app, phone number, address, or website. You can say, repeat or stop, at any time.

Speed up mechanism
• ignore the same questions (i.e. outputs) in different nodes
• not wait for every output reading
Speed up mechanism
• ignore the same questions (i.e. outputs) in different nodes
• not wait for every output reading

I: open c.s.n. premier collision
O: … To learn more, say, about us, services, mobile app, phone number, address, or website. You can say, repeat or stop, at any time.

I: yes
O: … To learn more, say, about us, services, mobile app, phone number, address, or website. You can say, repeat or stop, at any time.

I: about us
O: … Would you like to learn more?
Evaluation

- **28,904** skills from Amazon and **1,897** actions from Google through simulators

- Paths coverage rate: **90%**

- Error rate of answer generation:
  
<table>
<thead>
<tr>
<th>Yes/No</th>
<th>Instruction</th>
<th>Selection</th>
<th>Wh</th>
<th>Mix</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>8%</td>
<td>8%</td>
<td>5%</td>
<td>9%</td>
</tr>
</tbody>
</table>

- More than **5,200** hours spent. Each skill costs about 627 seconds on average

- The speed up mechanism saves **29.2%** time
Findings

Landscape

- Skills & authors
  - 68,066 skills from Amazon market with 12,376 different developer names
  - Some developers own more than 1,000 skills

- Invocation names
  - 9,799 skills’ invocation names do not meet Amazon’s requirements (e.g. using place/people names)
  - 2591 invocation names are not unique
Findings

Developer Specifications for request private information

a) using specific APIs (e.g., Alexa customer profile API) to request permissions

b) in the privacy policy of the skills
Findings

Developer Specifications for request private information

a) using specific APIs (e.g., Alexa customer profile API) to request permissions

b) in the privacy policy of the skills

Method

**Interactive records**

<table>
<thead>
<tr>
<th>WP</th>
<th>cop</th>
<th>nsubj</th>
<th>punct</th>
<th>nmod:poss</th>
<th>NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>What</td>
<td>is</td>
<td>your</td>
<td>name</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Collect verb (e.g. collect, use, …)**

**General term (e.g. personal data …)**

**Subsumptive relationships (e.g. such as, …)**

**Key components of the general declaration in PP**

- **Privacy Policy (PP)**
- **Permissions**
Findings

Developer Specifications for request private information

a) using specific APIs (e.g., Alexa customer profile API) to request permissions

b) in the privacy policy of the skills

Method

Name, phone number, …

interactive records

Results

• 1,141 skills conflict with the developer specifications
Findings

- **68 skills** have problems when users say “stop”
  - 32 skills change the default “stop” commands (e.g. I’ve done)
  - 29 skills ignore the stop command
  - 7 skills seem more strange
Conclusion

• Develop a **systematic** method to explore skills
  ➢ a suite of grammar-based approaches

• Conduct a **large scale** of testing in the skill market
  ➢ about 30,000 skills from markets

• Find **a good number** of skills that don't follow the development rules
  ➢ 1,141 skills request private information and 9,799 skills’ invocation names

• Find some suspicious skills
  • 68 skills have problems when receiving Stop command
Thanks for listening!

Q&A

guozhixiu@iie.ac.cn