Natural Language Understanding in Alexa

Alan Packer
October 2019
Agenda

• End-to-End SLU System Overview
• Core NLU System
• Entity Resolution
• Scale Challenges
• Deep Learning
• Rapid NLU Internationalization
• Q&A
Alexa Spoken Language Understanding – System Overview

User \(\xrightarrow{\text{Speech}}\) ASR \(\xrightarrow{\text{Text}}\) NLU \(\xrightarrow{\text{Labels}}\) Ranking / Dialog Manager \(\xrightarrow{\text{Actions}}\) TTS \(\xrightarrow{\text{Speech Output}}\)

**Clarification Dialog**

<table>
<thead>
<tr>
<th>Component</th>
<th>Input</th>
<th>Output</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automatic Speech Recognition (ASR)</td>
<td>Speech</td>
<td>Text (1-best or lattice of alternatives)</td>
<td>“Play Two Steps Behind by Def Leppard”</td>
</tr>
</tbody>
</table>
| Natural Language Understanding (NLU) | Text                | Slots and Intent Type         | Intent: PlayMusicIntent  
Slots:  
Artist Name=Def Leppard  
Song Title: Two Steps Behind |
| Ranking / Dialog Manager (DM)      | Labels & Context    | Dialog Actions                | Ask the application to play the song or clarify                          |
| Text-to-Speech (TTS)               | Text                | Speech                        | “Which artist?” or “Playing Two Steps Behind by Def Leppard”            |
Sidebar: A Search Engine

User → Query Understanding → Query Intents → Index Lookup / Answers → Candidates → Ranking → Actions → Result Rendering → HTML Output
Agenda

• End-to-End SLU System Overview
• Core NLU System
• Entity Resolution
• Scale Challenges
• Deep Learning
• Rapid NLU Internationalization
• Q&A
Statistical Natural Language Understanding (NLU)

**Utterance:** Play songs by CCR

**Entity Recognition:** CCR (ArtistName)

**Domain Classification:** Music

**Intent Classification:** PlayMusicIntent

**Entity Resolution:** Creedence Clearwater Revival

---

**Utterance:** Play songs by CCR

**Entity Recognition:** CCR (ArtistName)

**Domain Classification:** Music

**Intent Classification:** PlayMusicIntent

**Entity Resolution:** Creedence Clearwater Revival
Intent Classification (IC)

what’s the weather
tell me the forecast
will it rain today
please um tell me the weather for uh seattle
do I need an umbrella
what’s the expected low temperature today
Named Entity Recognition (NER)

- March twenty seventh
- A week from Sunday
- Eight days from now
- Easter Sunday
- Two days after Good Friday
Cross-Domain Re-ranking

<table>
<thead>
<tr>
<th>ASR Hyp</th>
<th>Intent</th>
<th>Dom Score</th>
<th>NER Score</th>
<th>IC Score</th>
<th>ASR Score</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>play frozen</td>
<td>PlayMusic</td>
<td>0.9</td>
<td>0.8</td>
<td>0.9</td>
<td>0.9</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>play frozen</td>
<td>PlayVideo</td>
<td>0.8</td>
<td>0.9</td>
<td>0.8</td>
<td>0.9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>play songs</td>
<td>PlayMusic</td>
<td>0.9</td>
<td>1.0</td>
<td>1.0</td>
<td>0.1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>...</td>
</tr>
</tbody>
</table>

Re-ranking Model → Re-ranker → Re-ranked hypotheses with scores

- **F1**: Device is connected to a screen
- **F2**: Detected slot type has an ER match
- **F3**: Detected entity exists in user’s library
- **F4**: Hyp. intent matches user’s past feedback

... ...
NLU Challenges

• Ambiguity in cross-domain intent classification
  – alexa, play frozen (Video.PlayVideoIntent or Music.PlayMusicIntent?)

• Difficulties in entity recognition
  – alexa, play dead mouse songs (ArtistName or AlbumName?)

• ASR errors
  – alexa, play dev anand music -> alexa, play david and music

• Rejecting out of domain utterances
  – alex, can you please take the trash out tonight?
Agenda

• End-to-End SLU System Overview
• Core NLU System
• Entity Resolution
• Scale Challenges
• Deep Learning
• Rapid NLU Internationalization
• Q&A
Entity Resolution (ER)

- Core NLU system produces *entity mentions*
- Mention must be linked to actual entity in application catalog
- For some slots, simply canonicalize the mention

<table>
<thead>
<tr>
<th>Entity Type</th>
<th>Mention</th>
<th>Entity Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ArtistName</td>
<td>kesha</td>
<td>Ke$hà (B002V5BVGE)</td>
</tr>
<tr>
<td>WeatherLocationCity</td>
<td>l. a.</td>
<td>Los Angeles</td>
</tr>
<tr>
<td>WeatherDate</td>
<td>tomorrow</td>
<td>2017-04-21</td>
</tr>
</tbody>
</table>
Entity Resolution Challenges

• User Queries in spoken form, catalog entries in written form
  – play songs by dead mouse -> Deadmau5

• Aliases, partial mentions, and ASR errors -> ambiguity
  – play trouble by taylor swift -> I Knew You Were Trouble

• Ranking and scoring is difficult, particularly across catalogs
  – play the frozen soundtrack
    • Frozen (Original Motional Picture Soundtrack) – Amazon Music
    • Frozen – IHeartRadio Album Catalog

• Internationalization and heterogeneous entity representations
Agenda

• End-to-End SLU System Overview
• Core NLU System
• Entity Resolution
• Scale Challenges
• Deep Learning
• Rapid NLU Internationalization
• Q&A
Scale challenges for Alexa NLU

- Rules vs. statistical models
  - Determinism/speed vs. generalization

- Ground truth data is expensive
  - Thousands of intents * dozens of locales
  - Measuring accuracy for 100K skills

- DNNs are expensive
  - Bigger, more complex models
  - Latency requirements
  - GPUs for training, inferencing?
Agenda

• End-to-End SLU System Overview
• Core NLU System
• Entity Resolution
• Scale Challenges
• Deep Learning
• Rapid NLU Internationalization
• Q&A
DNNs vs. linear models

• Baseline: 10% reduction in NLU error rate
  • Even bigger gains on the tail, low-resource domains or languages

• More importantly: Enables automation and scale
  • Unsupervised and Semi-Supervised learning
  • Active Learning and automated data selection
  • Few-shot / low-resource learning
  • Transfer Learning
  • Machine Translation
Transfer learning for NLU expansion

• Speed up natural language expansion for new domains
  • Increase accuracy of low resource domains

• Deep multitask architectures for NLU that leverage transfer learning from related tasks
  • Re-use existing annotated data from mature, high volume domains

• Evaluation on hundreds of new domains (internal and external developers)
  • Significant accuracy gains in low resource settings
  • Bootstrap functionality faster, with less data

“Fast and Scalable Expansion of Natural Language Understanding Functionality for Intelligent Agents”, Anuj Goyal, Angeliki Metallinou and Spyros Matsoukas, NAACL 2018
Transfer Learning

Source Labels

Source Model

Source data from mature domains (Weather, Music, Shopping etc.)

Transfer Learned Knowledge

Target Model

Target Labels

Data for the target domain, E.g. Traffic
Model Pre-training and Fine-tuning

• Pre-training: Build generic models on the source domains
  – Use data across tens of mature Alexa domains
  – Millions of annotated utterances
  – Hundreds of intents and slots

• Fine-tuning: Adapt models to the target domains
  – Few hundred or thousand annotated utterances for the target domain
  – Typically tens of intents and slots (depending on the domain)
Agenda

• End-to-End SLU System Overview
• Core NLU System
• Entity Resolution
• Scale Challenges
• Deep Learning
• Rapid NLU Internationalization
• Q&A
Bootstrap Lifecycle

- Golden Utterances
- Synthetic Data
- In House Data Collections
- Beta Traffic
- GA Live Traffic

Amount of Data

Amount of Annotated Data

Data Sparsity Bottleneck
Annotation Throughput Bottleneck
NLU i18n Bootstrap Scenarios

• Early Bootstrap: Low data scenario
  • Small amounts of data in new language
  • Large amounts of English annotated data
  • N-gram Projection
  • Full Sentence Translation

• Late Bootstrap: Incoming data but limited annotation resources
  • Large amounts of unannotated data in new language
  • Cross-lingual Annotation
Full Sentence Translation

“play music by katy perry” -> EN to DE Machine Translation -> “spiele music von katy perry” -> Build German NLU
Feature Level Translation (N-gram Projection)

1) Extract English N-gram Features
2) Project N-gram Features to German
3) Optionally embed projected features

English Training Data

"play music by katy perry"

spiele, spiele_musik, musik, musik_von

[0.13, -0.45, .......
[0.56, 0.12,........]

German Test Utterance

"spiel musik von katy perry"

spiel, spiel_musik, musik, musik_von

[0.12, -0.41, .......
[0.56, 0.12,.........]
Cross-Lingual Translation

“spiele musique von katy perry” → DE → EN Machine Translation → “play music by katy perry” → English NLU System → Project Annotation EN → DE → PlayMusicIntent ArtistName: katy perry

PlayMusicIntent spiele musique von katy|ArtistName perry|ArtistName
Agenda

- End-to-End SLU System Overview
- Core NLU System
- Entity Resolution
- Scale Challenges
- Deep Learning
- Rapid NLU Internationalization
- Q&A