Generalized Zero-shot Intent Detection via Commonsense Knowledge

A.B. Siddique, Fuad Jamour, Luxun Xu, and Vagelis Hristidis
{msidd005, fuadj, luxu051}@ucr.edu, vagelis@cs.ucr.edu

Motivation
- Identifying user intents from natural language utterances is crucial in conversational systems.
- It has been extensively studied as a supervised classification problem.
- However, in practice, unseen intents emerge after deploying the model and they do not have any training data.
- We propose RIDE: a generalized zero-shot intent detection model that seamlessly adapts and classifies natural language utterances with both seen and unseen intents.
- RIDE computes robust and generalizable relationship meta-features that capture deep semantic relationships between utterances and intent labels.

These meta-features are computed by considering how the concepts in an utterance are linked to those in an intent label via commonsense knowledge.

Computation of Relationship Meta-features

User Utterance: I am feeling hungry.
Intent Label: FindRestaurant
Relationship Meta-features: (look for, Synonym, find)

Evaluation
- Main result: F1 scores for competing models in the generalized zero-shot setting.

<table>
<thead>
<tr>
<th>Method</th>
<th>SNIPS Unseen</th>
<th>SNIPS Seen</th>
<th>SGD Unseen</th>
<th>SGD Seen</th>
<th>MultiWOZ Unseen</th>
<th>MultiWOZ Seen</th>
</tr>
</thead>
<tbody>
<tr>
<td>BERT-SPC</td>
<td>0.2761</td>
<td>0.7152</td>
<td>0.1872</td>
<td>0.6401</td>
<td>0.1932</td>
<td>0.6413</td>
</tr>
<tr>
<td>IntentCapsNet</td>
<td>0.0000</td>
<td>0.6332</td>
<td>0.0000</td>
<td>0.5508</td>
<td>0.0000</td>
<td>0.6038</td>
</tr>
<tr>
<td>ReCapsNet</td>
<td>0.1601</td>
<td>0.6783</td>
<td>0.1331</td>
<td>0.5751</td>
<td>0.1467</td>
<td>0.6170</td>
</tr>
<tr>
<td>SEG</td>
<td>0.6991</td>
<td>0.8651</td>
<td>0.4032</td>
<td>0.6356</td>
<td>0.4143</td>
<td>0.6456</td>
</tr>
<tr>
<td>RIDE w/o PU</td>
<td>0.9103</td>
<td>0.8799</td>
<td>0.4634</td>
<td>0.8295</td>
<td>0.4645</td>
<td>0.8816</td>
</tr>
<tr>
<td>RIDE w/ PU</td>
<td>0.9254</td>
<td>0.9080</td>
<td>0.5734</td>
<td>0.8298</td>
<td>0.5206</td>
<td>0.8847</td>
</tr>
</tbody>
</table>

RIDE Overview
- LSTM
- RMG
- Prediction Function P(T_j | x_i)

Table 3. Utilizing utterance and intent embeddings only (i.e., UI-Embed) results in very low F1 score, i.e., 23.67% on SNIPS. Employing relationship meta-features only (i.e., Rel-M) results in significantly improving the results of all the competing models. For SNIPS, the role of the PU classifier is incorporated or not. For SNIPS, the role of the PU classifier is significant.

ACKNOWLEDGMENTS
Interested readers can refer to [29] for more details about this work.

Insertion of generation 15%.

Figure 2: F1 scores for unseen intents for the competing models. For both seen and unseen intents, our model outperforms the SOTA method. For SNIPS, the role of the PU classifier is significant.

F1 scores of all models with and without PU classifier are shown in Table 3. For unseen intents, our model significantly improves the results of all the competing models. For SNIPS, the role of the PU classifier is significant.

Figure 2: F1 scores for unseen intents for the competing models. For both seen and unseen intents, our model outperforms the SOTA method. For SNIPS, the role of the PU classifier is significant.

F1 scores of all models with and without PU classifier are shown in Table 3. For unseen intents, our model significantly improves the results of all the competing models. For SNIPS, the role of the PU classifier is significant.