Overview

Task: Use active learning to minimise training data for deep semantic parsers.

Challenge 1: How to apply active learning to “overnight” data collection?

Challenge 2: How to set hyperparameters without full training data?

“Overnight” Data Collection

1. Generate logical form from stochastic grammar
   \[
   \text{argmax(type.article, publicationDate)}
   \]

2. Translate to “clumsy” prompt
   “article that has the largest publication date”

3. Crowd workers produce fluent paraphrases
   “what is the newest published article?”

Overnight Active Learning

Hyper Tuning

<table>
<thead>
<tr>
<th>Configuration From</th>
<th>NLMap</th>
<th>Social</th>
<th>ATIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATIS</td>
<td>76.0</td>
<td>65.8</td>
<td>86.0</td>
</tr>
<tr>
<td>20% dataset</td>
<td>84.2</td>
<td>68.9</td>
<td>85.7</td>
</tr>
<tr>
<td>Full dataset</td>
<td>84.2</td>
<td>69.1</td>
<td>86.0</td>
</tr>
<tr>
<td>SOTA</td>
<td>84.1</td>
<td>68.8</td>
<td>86.1</td>
</tr>
</tbody>
</table>

Forward S2S - Least Confidence Score

\[
x' = \arg\min_{x \in U_x} \left[ \max_{y^*} P(y^* | x; \theta) \right]
\]

P is computed by a Seq2Seq model with attention, requires utterance \(x\) but not logical form \(y\).

Backward Classifier

- Active learning score = linear combination of features using weights from binary classifier.
  - Predict if Forward S2S selects utterances.
  - Trained on ATIS dev corpus.
- Binary classifier to predict Forward S2S using
  - RNN LF language model
- Backward S2S model
  - Margins between the best and second best hypotheses
  - Source token frequency
  - Utterance log loss
  - Encoder and decoder last hidden states

Conclusions

- 20% (10% dev + 10% train) of the full dataset is sufficient for hyperparameter tuning.
- Least confidence forward active learning score doesn’t apply to “overnight” collection.
- Either backward S2S or classifier scores work on all corpora.