Active Learning for Deep Semantic Parsing

Long Duong, Hadi Afshar, Dominique Estival, Glen Pink, Philip Cohen, Mark Johnson

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
   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
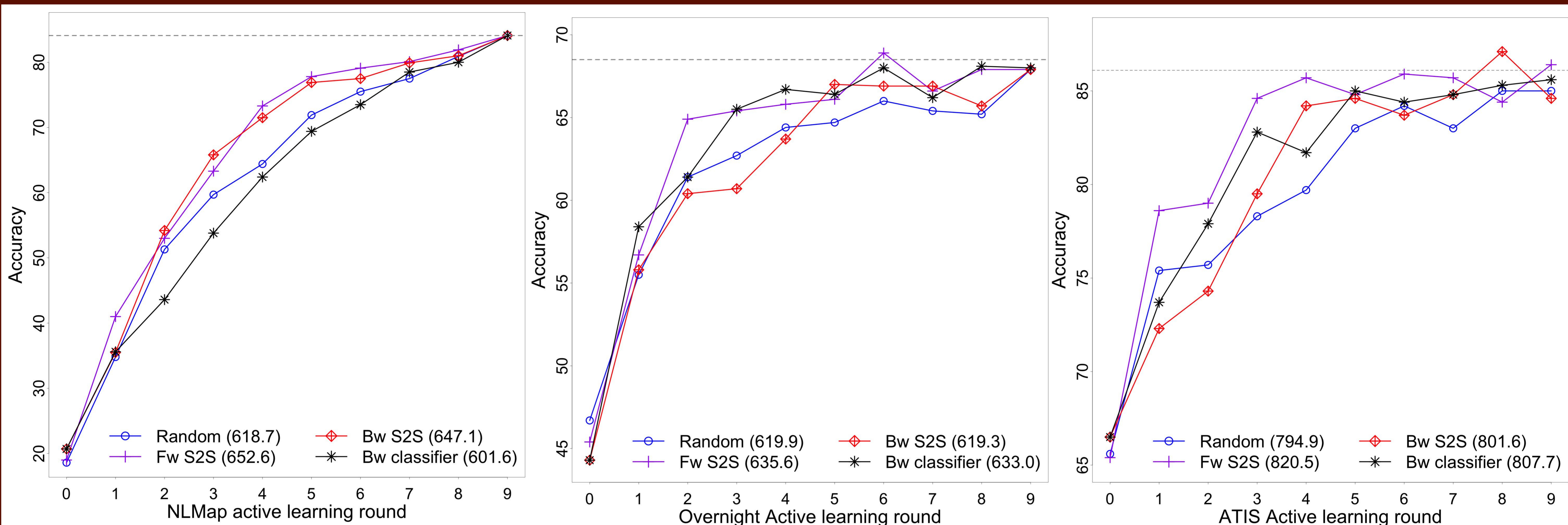
Hyper Tuning

Configuration From	NLMap	Social	ATIS
ATIS	76.0	65.8	86.0
20% dataset	84.2	68.9	85.7
Full dataset	84.2	69.1	86.0
SOTA	84.1	68.8	86.1

Forward S2S - Least Confidence Score

$$x' = \operatorname*{argmin}_{x \in U_x} \left[\max_{y^*} \mathsf{P}(y^*|x;\theta) \right]$$

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

Backward S2S - Least Confidence Score

$$y' = \operatorname*{argmin}_{y \in U_y} \left[\max_{x^*} \mathsf{P}(x^*|y;\theta) \right]$$

y : query(nwr(keyval('craft','distillery')),qtype(count))
x_1 : How many distilleries do you count?
x_2 : How many distilleries are there?
x_3 : Tell me the number of distilleries.

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.

Reference: Wang et al. (2015) "Building a Semantic Parser Overnight"