Compositional pre-training for neural semantic parsing

Amir Ziai (amirziai@stanford.edu)

Motivation

Semantic parsing is the task of converting **natural language utterances** into **machine-executable logical forms**. Examples of this parsing include asking questions that are then converted to queries against a database, generating code from natural language, and converting natural language instructions to an instruction set that can be followed by a system.

Traditionally this task has been tackled by a combination of heuristics and search. These systems were very complex and brittle. The majority of the research has shifted to using neural systems.

Problem

Constrained decoding

Decoded sequence need to be **constrained** by what would constitute a valid logical form. Similar to other conventional seq-2-seq use caess a single source sentence can have multiple valid target sequences.

Prior knowledge

Logical forms are highly structured and a vanilla seq-2-seq model does not allow for injecting prior knowledge.

Dataset

We used the GeoQuery dataset. This dataset consists of a set of questions about US geography facts and the corresponding Prolog query which represents the logical form. We use 600 examples for training and 280 examples for testing. The task is to produce the logical form given the question.

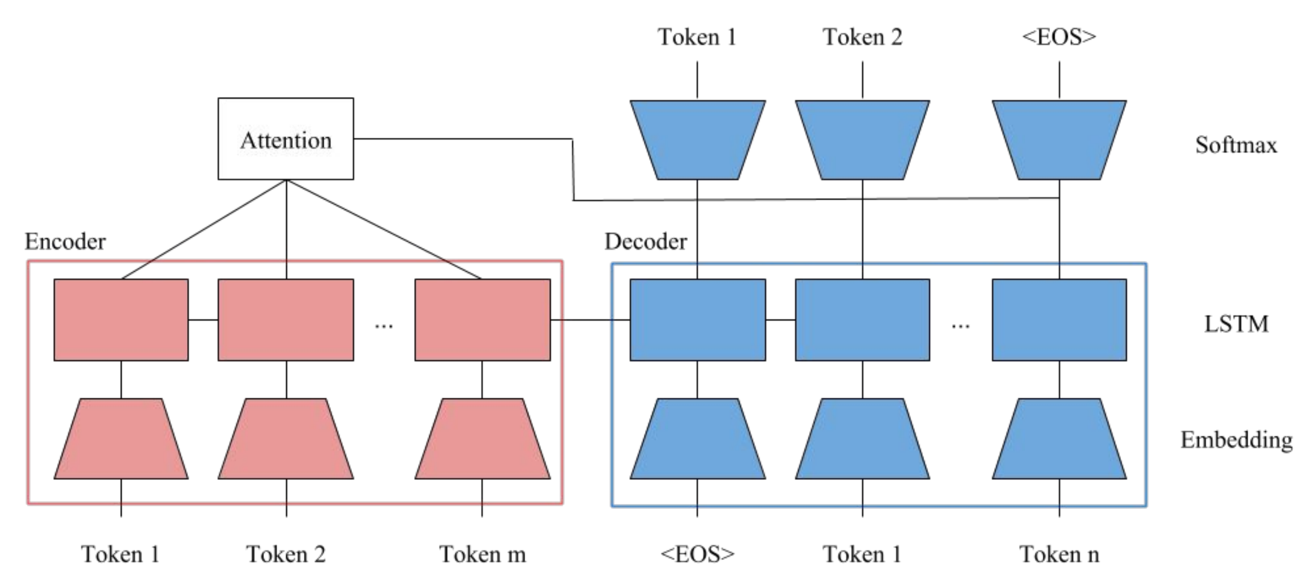
Example training pair:

```
- What is the capital of Alaska?
- _answer(NV,(_capital(V0),_loc(V0,N
  V),_const(V0,_stateid(alaska))))
```

Approach

Architecture

Using a seq-2-seq architecture with a copying attention mechanism. Encoder is a bi-LSTM with 256 hidden units, tokens embedded size is 64, and an Adam optimizer with 0.001 learning rate is used.



Evaluation metric

Sequence accuracy is used for quantifying the quality of the generated logical forms which is defined as an exact string match between the true and predicted sequence. This metric can be overly strict and therefore we also use token accuracy, which is the true tokens which are generated by the model.

Data augmentation

Three augmentation strategies are used from [1]. The first strategy abstracts an entity with its type. The second strategy looks for matches between input and output and replace both with the entity type. The third strategy concatenates the sequence k times (k=2 in this work).

Co-occurrence augmentation

This novel strategy links tokens that co-occur in similar contexts with some minimum support and randomly replaces associated tokens from a generated lookup table. For example we can generate 49 augmented training pairs by substituting Ohio for other US states in this example:

```
What is the highest point in Ohio?
```

Results

Pre-training and data augmentation

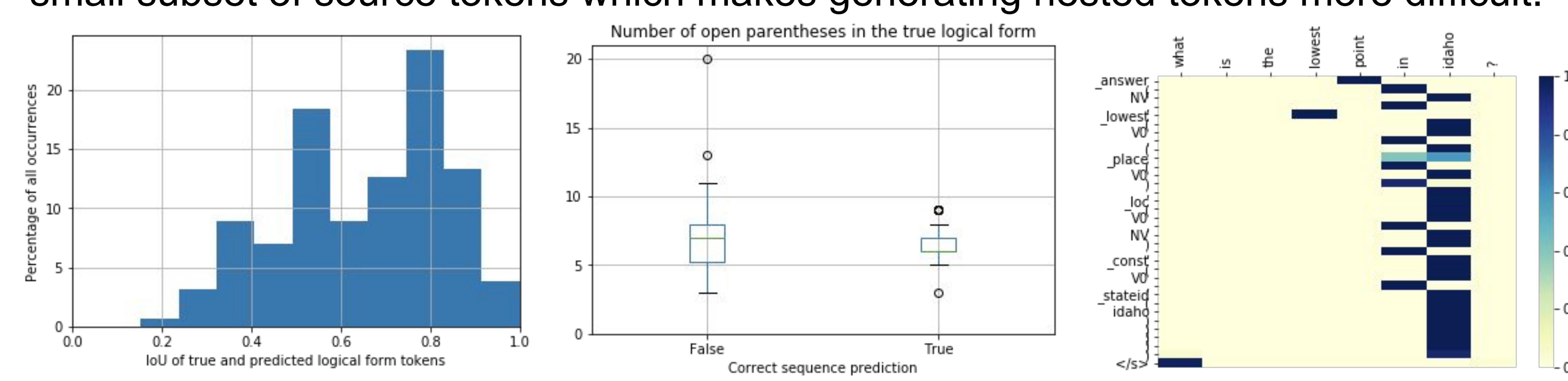
Inspired by recent success stories such as BERT[2] we experimented with pre-training the model with augmented examples and fine-tuning the training on a given task.

Augmentation	Pre-train	Sequence accuracy	Token accuracy
nesting, entity, concat	Yes	**74.3%**	87.8%
nesting, entity, concat	No	73.6%	**88.2%**
nesting, entity, concat, co-occurrence	Yes	66.1%	87.3%
nesting, entity, concat, co-occurrence	No	56.1%	81.7%

Pre-training on augmented examples followed by fine-tuning results in the highest sequence accuracy.

Analysis

We conducted error analysis to gain insight into the model. Figure (a) shows the Intersection over Union (IoU) of the unique tokens in the predicted and true logical forms (for incorrect examples). This graph suggests that most tokens are predicted by they may not have the correct nesting or order. Figure (b) shows the that the model fails as the nesting depth of the logical form (using the number of opening parenthesis as a proxy) increases. Finally Figure (c) shows that attention tends to be focused on a small subset of source tokens which makes generating nested tokens more difficult.



(a) (b) (c)

Conclusions

Pre-training with augmented examples can be beneficial for semantic parsing. With this framework we can use monolingual unsupervised training of decoder and the encoder and use more augmentation strategies.

References

[1] Robin Jia and Percy Liang. Data recombination for neural semantic parsing. arXiv preprint arXiv:1606.03622, 2016.

[2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.