Column Attention in SQLNet Spider-Task Architecture

Davey Proctor and Dragomir Radev

Department of Computer Science, Yale University

LILY Lab

Introduction

The Spider dataset is the largest text to complex SQL dataset yet. In addition to complex SQL syntax, the dataset additionally features databases across multiple domains, forcing models to generalize beyond simply optimizing for semantic parsing tasks on a specific subject. Its complexity is represented in transcending "SELECT…WHERE..." style queries; instead using "GROUP BY," nested, "HAVING", and "JOIN"-ing over multiple tables.

Several models have been proposed for addressing this new semantic parsing task. One of the most promising baselines is adapted from SQLNet, a state-of-the-art model for the WikiSQL task. This model was initially used for WikiSQL's more simply-structured queries, but it has been extended to capture the greater complexity in the Spider task.

My contribution is to experiment with the adapted SQLNet model and expose its effectiveness on Spider, while also suggesting improvements.

Method

SQLNet employs a sketch approach that corresponds to the SQL grammar and treats SQL generation as a slot-filling exercise within this sketch. This improves over other models in that it fundamentally avoids sequence-to-sequence structure when order does not matter, such as each column to be conditioned over in a multiple "WHERE" clause (see Figure 1).

One of the major insights of the SQLNet model was the use of column attention. Rather than directly proceeding with the question and column encoding, the question is re-encoded conditional on each column. Intuitively, if the model must decide whether to fill a certain column into a slot, it is difficult to directly use the question encoding, which might not remember important word-level information related to the correct column. Instead, with column attention, the model can upweight words in the question most relevant to classification tasks specific to that column.

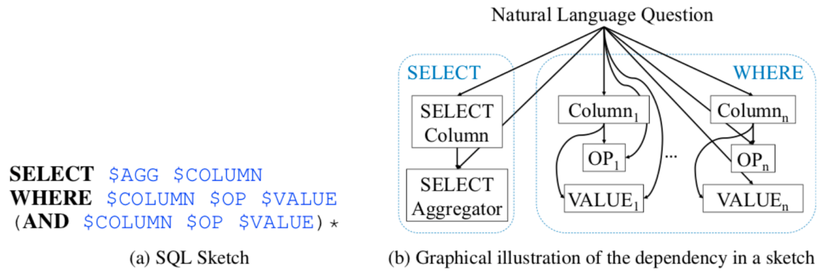

Figure 1. SQLNet Sketch-Based Model

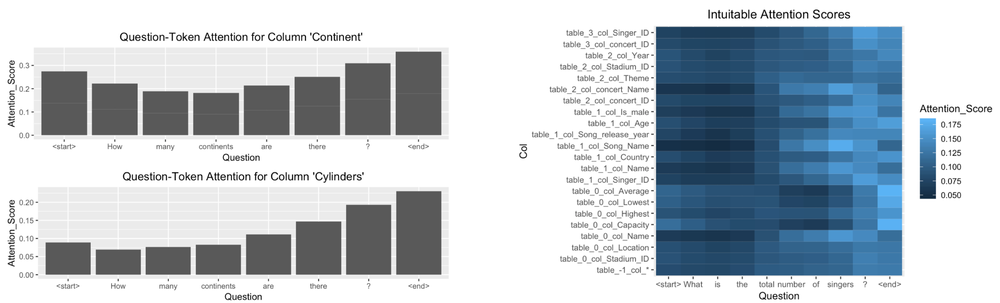

Figure 2. Column Attention

Column Attention?	Select	Where	Group By	Order By	Total	Runtime
Yes	30.2	54.6	73.2	85.5	10.4	00:52:46
No	30.3	51.5	72.0	86.0	10.2	00:45:55

Figure 3. Spider-task Dev Scores for SQLNet Model with and without Column Attention

Results

My experiments centered on the problem of understanding the column attention mechanism, specifically when applied to the Spider task.
First, The SQLNet paper reports a 3 point improvement when they use column attention; I find no similar improvement in the adapted SQLNet model applied to Spider (see Figure 3).

Second, although the SQLNet paper emphasizes "column attention is a special instance of the generic attention mechanism," they make no attempt to visualize this attention as has become popular in other settings where attention has been used.

Thus, I sought to visualize the attention mechanism, shown in Figure 2. Although somewhat intuitable, we notice that the attention still seems to be slightly biased towards the start and end of the sentence, implying that the encoding of the bi-LSTM at the start and end might by-and-large be sufficient for the task.

Conclusion

Although the SQLNet model has been reapplied with some success to the Spider task, there is clearly room for improvement. Towards this end, I sought to experiment with what gain the column attention is giving the new model on the Spider task.

On the Spider task there is hardly an improvement for using attention. Attention lends interpretability to the model through the attention plot, but suggests better-honing the model could yield better results. I propose a future direction of self-attention when learning the column embeddings, since fundamentally order of columns in the table does not matter. Furthermore, in the paper "Attention is All you Need," the authors find that attention can improve over bi-LSTM, which is currently used to represent column meaning.

Acknowledgement

Tao Yu and Rui Zhang helped set up SQLNet in a format that was easy to experiment with, and they also helped me brainstorm ideas of what experiments to do and what future directions would be worth-while.