Sequence-Aware Query Recommendation Using Deep Learning

Eugenie Y. Lai | Advised by Rachel Pottinger

University of British Columbia, Data Management and Mining Lab

BACKGROUND & MOTIVATION

Objective: help users write SQL queries using past queries

- Database management systems understand SQL queries
- **SQL queries** are questions in structured query language
- **Challenge**: users lack field- & database-related expertise

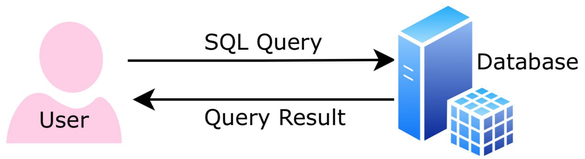

Figure 1: User interaction with databases.

Intuition: predict the user's next query by learning from the queries posted by past users

PROBLEM & PRELIMINARIES

Model query recommendation as a **query prediction** task

```
SELECT j.target, CAST(j.estimate AS VARCHAR) AS estimate
FROM Jobs j, Status s,
        (SELECT DISTINCT target, queue FROM Servers r
         WHERE r.queue NOT IN (SELECT MIN(queue)
                               FROM Servers
                               GROUP BY target))
WHERE j.outputtype LIKE '%QUERY%'
```

Figure 2: Sample SQL query Q.

Use **sequence-to-sequence** (seq2seq) models

- Advantages: less human intervention, etc.
- Other applications: sentence-level NLP (e.g., chatbot)
- Recurrent neural networks (RNNs)

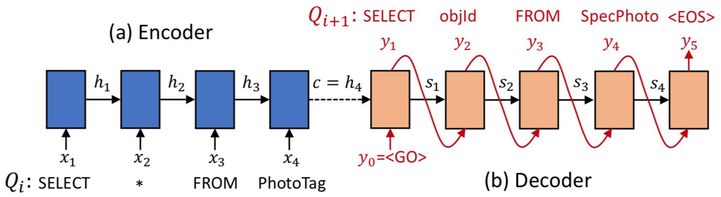

Figure 4: An RNN seq2seq model that takes query Q in a session and predicts the next query Q*.

CONTRIBUTIONS

Leverage **whole queries** and **query session sequences**

- Define a new approach to guide DBMS users' next-step query formulation
- Adapt a broad set of deep learning models to our problem
- Empirically evaluate our approach using two real-world datasets and compare to the existing approach

METHOD OVERVIEW

First, train seq2seq models

- **sequence-aware**: with query prediction task using query subsequences $\langle Q_i, Q_{i+1} \rangle$.
- **sequence-blind** (in comparison) with query reconstruction task using $\langle Q_i, Q_i \rangle$.

Then, recommend query fragments using the trained model

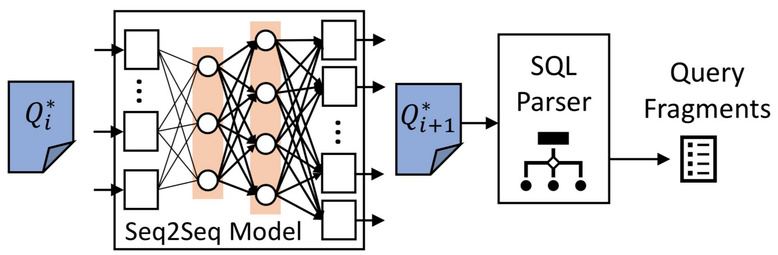

Figure 5: Query fragment prediction.

EXPERIMENTAL SETUP

Evaluate the efficacy of the **combination** of deep learning models and query subsequences in query recommendation

- **Task**: Use test set, given Q_i, predict **fragments** in Q_{i+1}
- **Methods compared**
  - baseline1: the most popular queries
  - QueRIE framework: existing method
  - seq-aware vs. seq-blind RNN seq2seq models
- Metric: F-measure

RESULTS & DISCUSSION

Fragment prediction result shows that **seq-aware** RNN

- **outperforms** other approaches by far in **table & function**
- slightly outperforms others in **attribute** prediction

The seq-blind RNN performs best in literal prediction

- may suggest weak sequential patterns in changes in literal

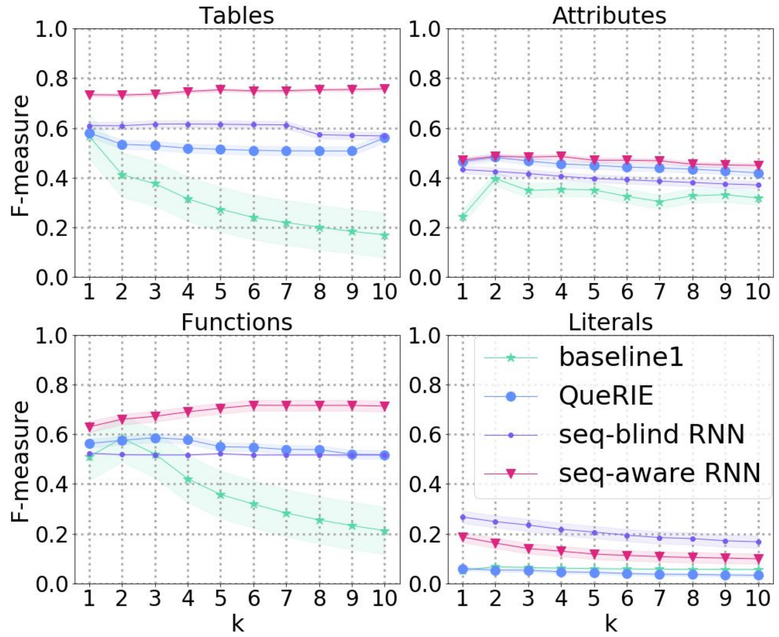

Figure 6: ***k*** is the number of model-predicted queries. Shadow is the 95% confidence interval, by sampling 5% test data 10 times.

CONCLUSION & FUTURE WORK

Deep learning + query session sequences is effective

Next steps: strengthen the evaluation

- Conduct a user study
- Compare to more existing methods
- Evaluate time complexity
- Evaluate the semantic distance of query fragments

Contact: https://eugenielai.github.io/