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

Help users write structured query language (SQL) queries.

- **Database management systems (DBMS)**:
  - Provide an infrastructure for high volumes of data
  - Understand SQL queries
- **User's challenge**: field- and database-related expertise
- **Query recommendation**: facilitate user query formation

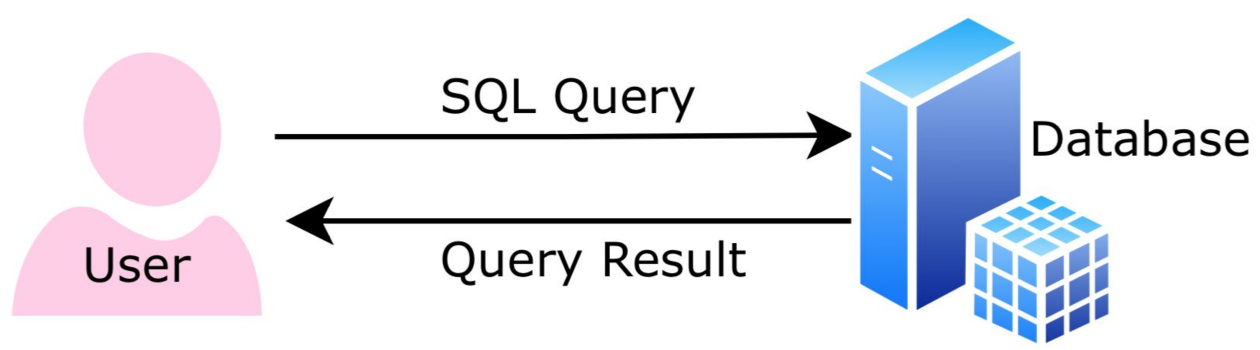


Figure 1: User interaction with databases.

- **Two limitations** of the existing approaches
  - Ignore the nature SQL query session sequences
  - Rely on human-selected features

PROBLEM DEFINITION

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

- Query fragment: tables, attributes, functions, literals
- Query template: consist of SQL keywords

```
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.

```
SELECT Attribute, Function(Attribute AS VARCHAR)
FROM Table, Table,
  (SELECT DISTINCT Attribute, Attribute FROM Table
   WHERE Attribute NOT IN (SELECT Function(Attribute)
                              FROM Table
                              GROUP BY Attribute))
WHERE Attribute LIKE Literal
```

Figure 3: Query template of 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

SEQUENCE-TO-SEQUENCE MODELS

Use sequence-to-sequence (seq2seq) models.

- A deep learning architecture
- **Advantages**: less human intervention, inductive biases catered to queries, capacity to leverage large data
- Closely related field: **sentence-level NLP** (e.g., chatbot)
- **Models**: recurrent neural networks (RNNs), transformer, convolutional neural networks (CNNs)

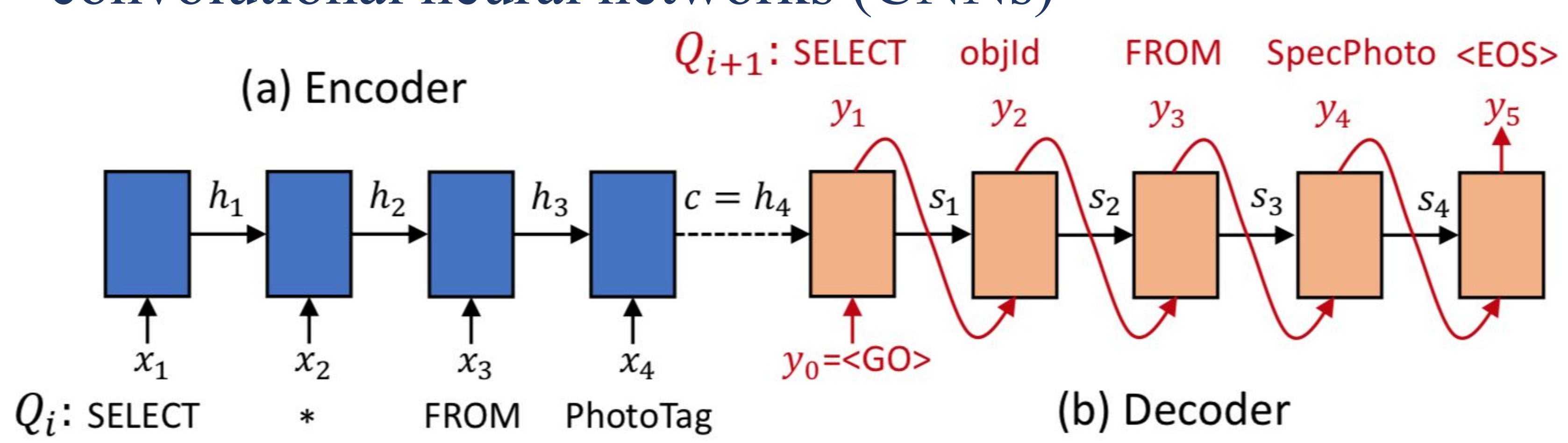


Figure 4: An RNN seq2seq model that takes query Q in a session and predicts the next query Q^*.

METHOD OVERVIEW

Build **sequence-aware** seq2seq models.

- Train on query prediction task using query subsequences: pairs of queries $\langle Q_i, Q_{i+1} \rangle$
  - Model input sequence: Q_i
  - Model target sequence: Q_{i+1}
- Build vocabulary at word level, replace numeric values

Build **sequence-blind** seq2seq models in comparison.

- Train on query reconstruction task using $\langle Q_i, Q_i \rangle$

METHOD OVERVIEW (cont.)

Recommend query fragments using the trained model.

- Given a user input Q_i^*, the model predicts next-step query
- Then our SQL parser gets predicted query fragments

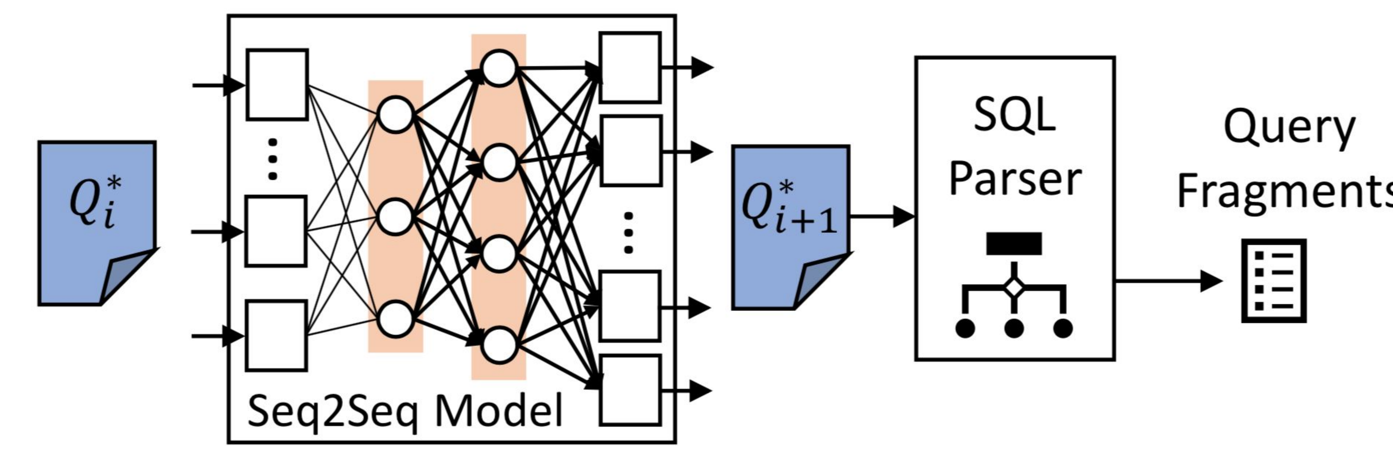
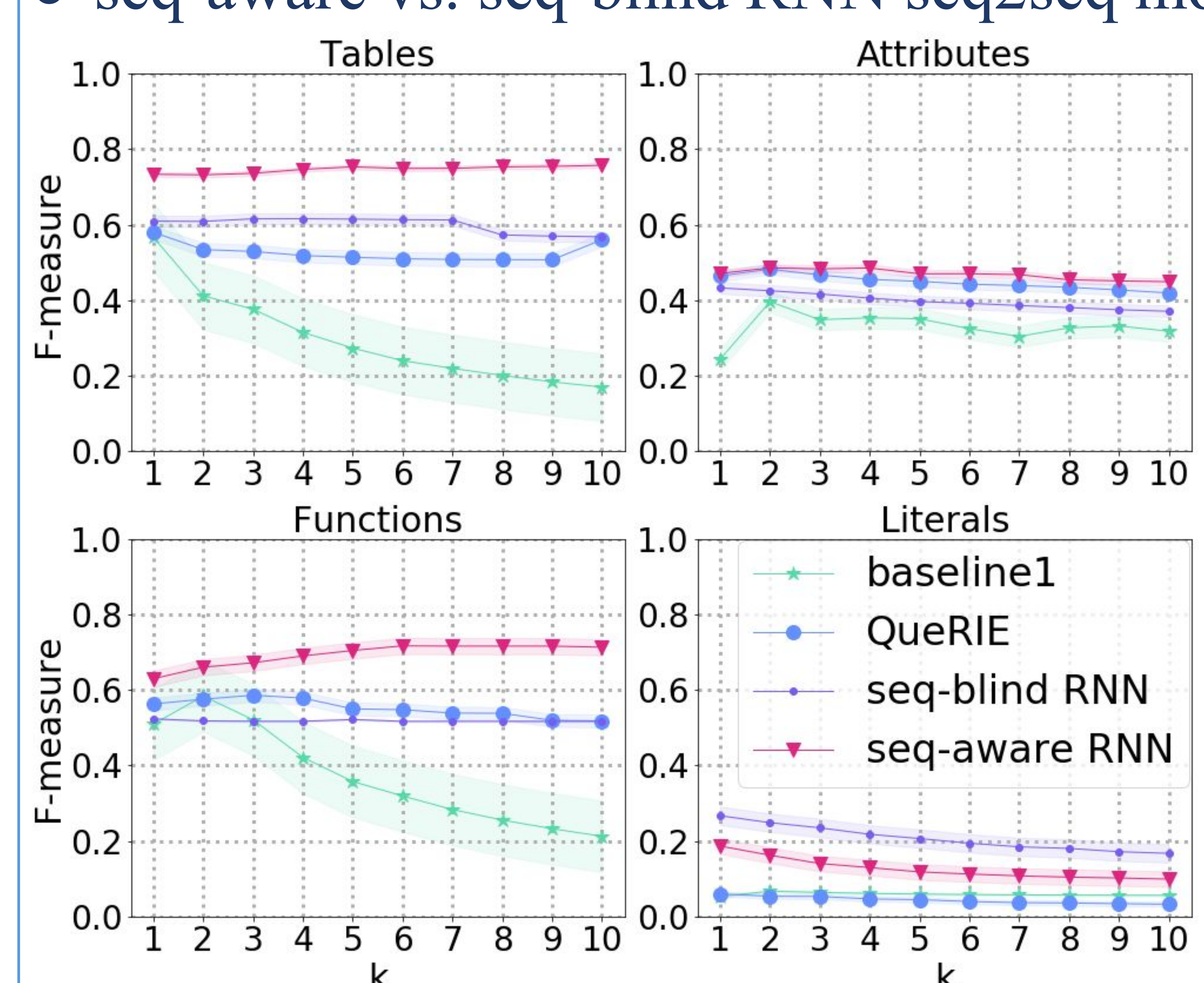


Figure 5: Query fragment prediction.

EXPERIMENTAL EVALUATION & DISCUSSION

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

- baseline1: the most popular queries
- QueRIE framework: existing method
- seq-aware vs. seq-blind RNN seq2seq models



Task: In test set, given Q_i, predict fragments in Q_{i+1}.

Figure 6: Fragment prediction result shows that seq-aware models outperform other approaches. ***k*** is the number of model-predicted queries. Shadow is the 95% confidence interval.

CONCLUSION & FUTURE WORK

Deep learning + query session sequences is effective.

Next steps: include more session info, an user study, etc.