ICDE 2023
Anaheim, CA April 3 – 7, 2023

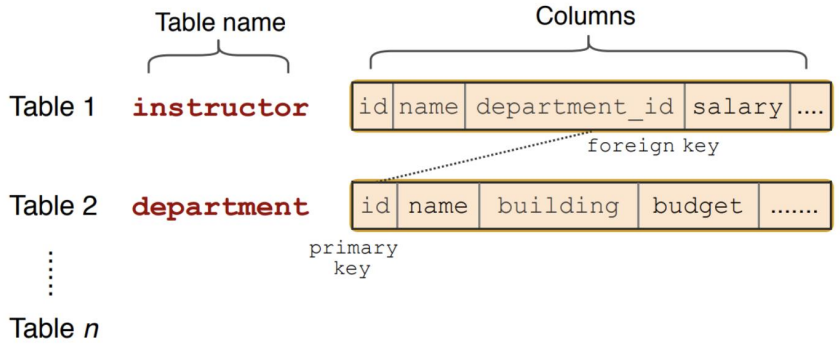
GAR: A Generate-and-Rank Approach for Natural Language to SQL Translation

*Yuankai Fan, Zhenying He, Tonghui Ren, Dianjun Guo, Lin Chen, Ruisi Zhu, Guanduo Chen, Yinan Jing, Kai Zhang, X.Sean Wang

Fudan University

What is NL2SQL?

Database Schema



+

NL Query

Complex question

What are the name and budget of the departments with average instructor salary greater than the overall average?

⬇

SQL Query

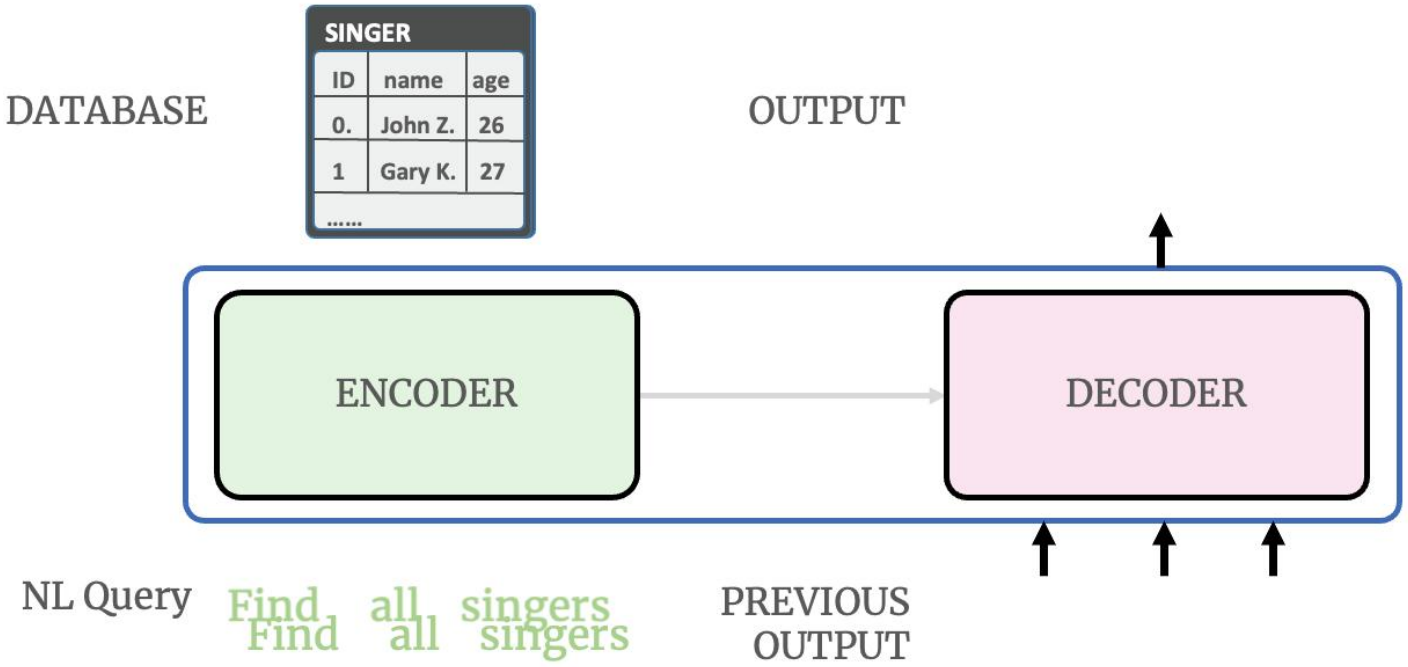
Complex SQL

```
SELECT T2.name, T2.budget
FROM instructor as T1 JOIN department as
T2 ON T1.department_id = T2.id
GROUP BY T1.department_id
HAVING avg(T1.salary) >
    (SELECT avg(salary) FROM instructor)
```

Spider example

Mainstream Approach

- Based on **Sequence-to-sequence** framework
- Based on **pre-trained language models**



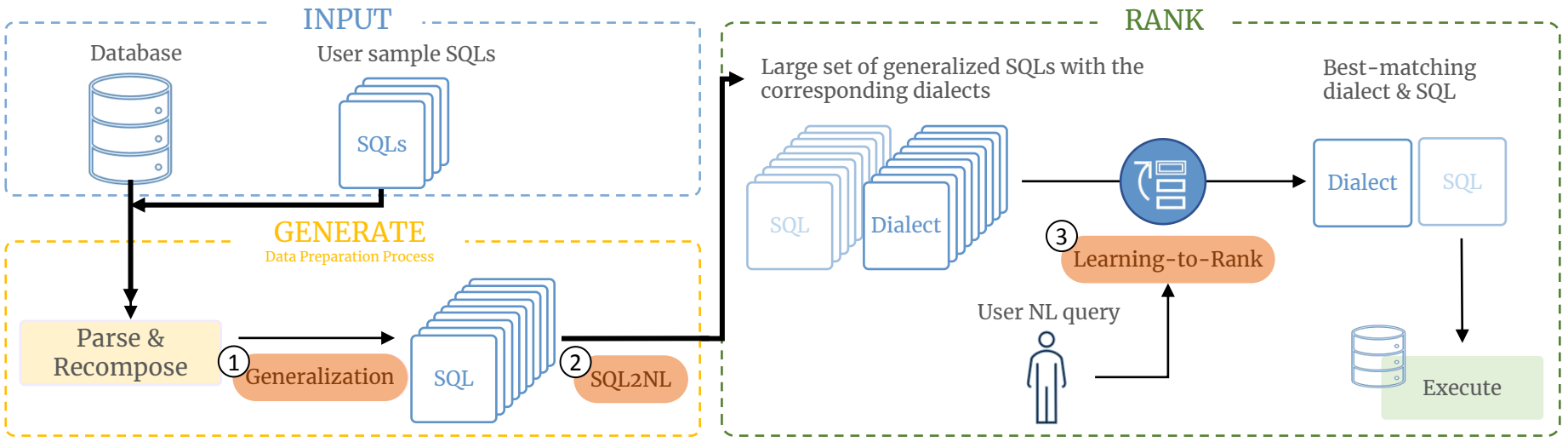
Existing Problems

- Most erroneous translations happen with **complex queries**
- Language models have limited understanding of the structure and semantics specific to a database

Model	Easy	Medium	Hard	Extra Hard	Overall
GAP	0.915	0.742	0.644	0.494	0.727
SMBOP	0.890	0.791	0.644	0.470	0.737

What is GAR?

- GAR: Generate-And-Rank
- Start with a set of sample queries
  1. **Generalization**
  2. **SQL2NL**
  3. **Learning-to-rank**



Generalization

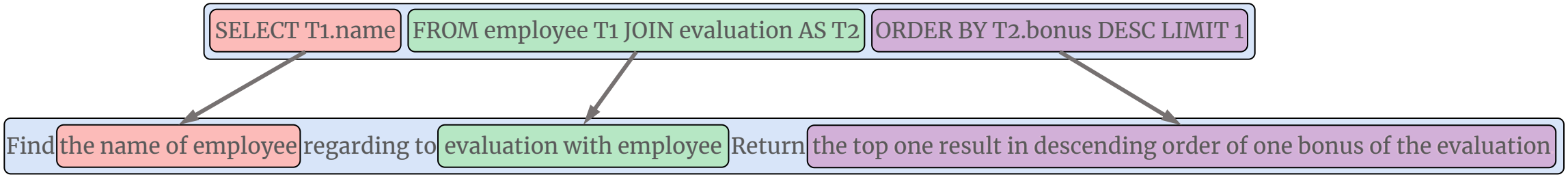
- **Decompose-then-recompose**
  1) Decompose into a set of ***query components***

Generalization

- **Decompose-then-recompose**
  2) Recompose the components to generalize

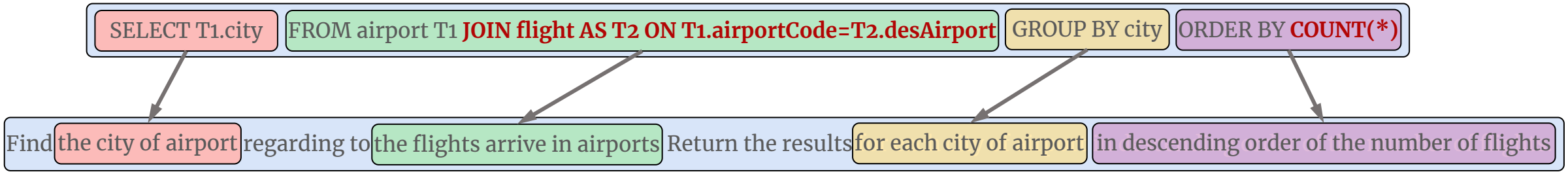
SQL2NL

- **Goal**: SQL semantics space → NL semantics space
- **Template-based** SQL2NL approach
  - ➢ Less natural NL expression (***Dialect expression***)



SQL2NL with JOIN semantics

- JOIN operations may ***raise the level of abstraction***
- JOIN semantics
  - ✓ *Join tables*
  - ✓ *Join condition*
  - ✓ *Join description*
  - ✓ *Table keys*



Learning-to-Rank

- Rank based on *semantic similarity*
- **Two-stage ranking pipeline**
  1) Coare-grained retrieval model
  2) Fine-grained re-ranking model

Evaluation

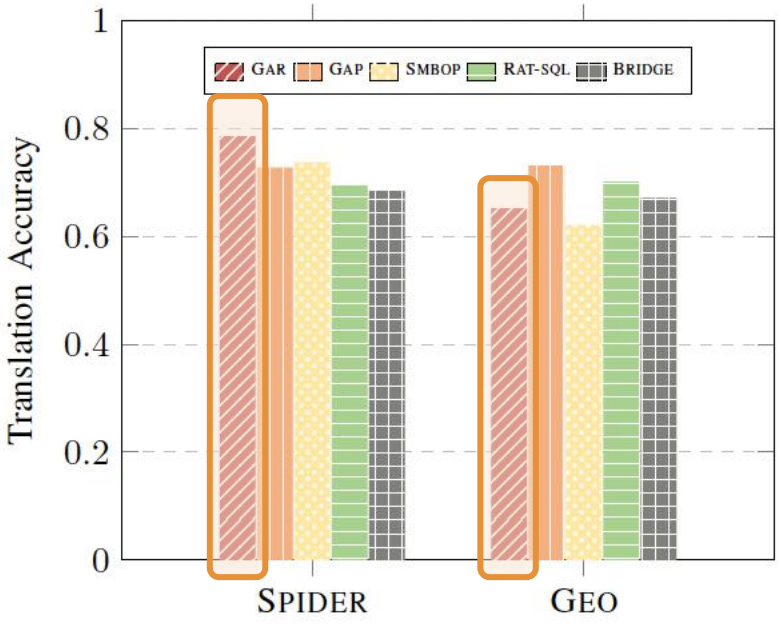
- Benchmarks
  - ✓ Spider
  - ✓ GEO
  - ✓ MT-TEQL
  - ✓QBEN *(our work)*

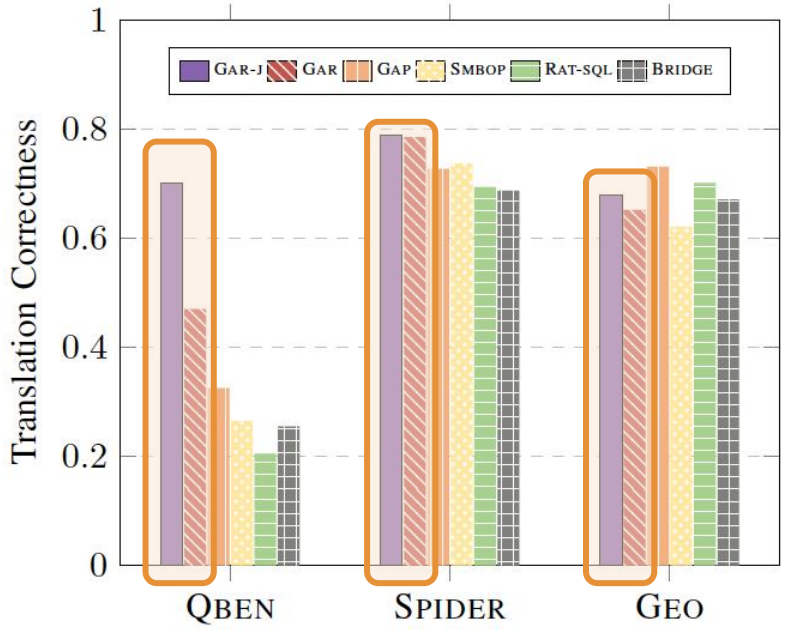
- Metrics
  - ✓ Translation Accuracy *(syntactic equivalence)*
  - ✓ Execution Accuracy
  - ✓ Translation Precision/MRR *(for ranking evaluation)*

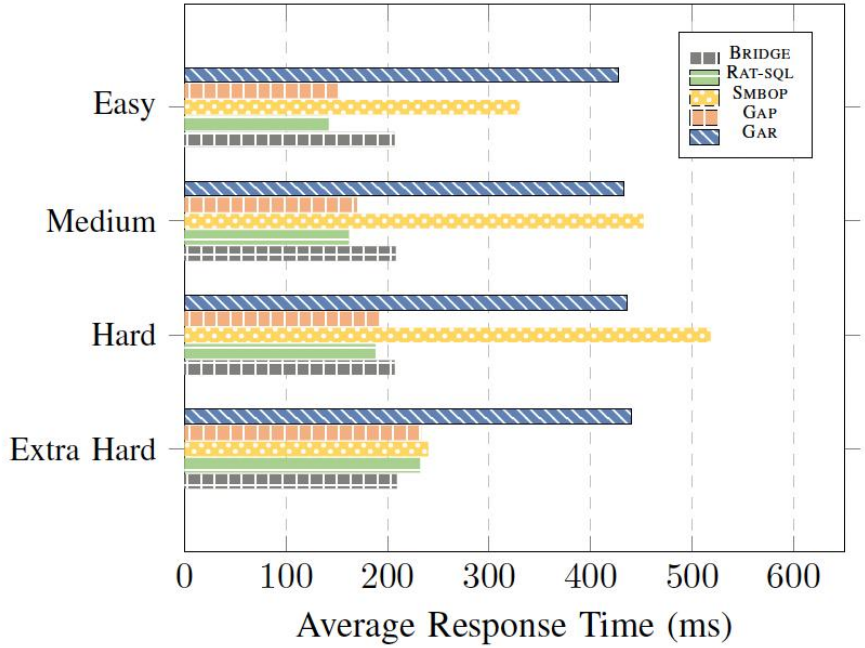
Evaluation

- Baselines
  - *RAT-SQL*
  - *GAP*
  - *SmBoP*
  - *BRIDGE*
- Sample query selection
  - Use validation/test set to generalize and then rule out all the ground truth queries

Evaluation







Model	Overall	*Exec.*
GAR + SPIDER validation set	0.784	0.693
SMBOP	0.726	0.705
BRIDGE	0.648	0.626
GAP	N/A	N/A
RAT-SQL	N/A	N/A

MT-TEQL

Dataset	MRR	Precision@1	Precision@3	Precision@10
SPIDER	0.823	0.785	0.859	0.875
GEO	0.680	0.652	0.679	0.680

Conclusion

- Sample queries bring database semantics
- GAR: A generative approach for NL2SQL
  - **Generate-then-rank > Sequence-to-sequence!**

Yuankai Fan <ykfan19@fudan.edu.cn>

https://github.com/Kaimary/GAR