Query-based Workload Forecasting for Self-Driving Database Management Systems

Lin Ma, Dana Van Aken, Ahmed Hefny, Gustavo Mezerhane, Andrew Pavlo, Geoffrey J. Gordon
Self-Driving Cars

Perception

Actions

Planning
First Step Towards Self-Driving Databases

Workload Forecasting

Indexing

Partitioning

Scaling
First Step Towards Self-Driving Databases

Workload Forecasting

Time

Planning

Observation
Workload Forecasting

• When, how many, and what queries will arrive

Goals:
1. Good Accuracy
2. Major Patterns
3. Cost vs. Accuracy

Prediction Horizon
Prediction Interval
Challenges

• Support for dynamic workloads
• Support large query volumes
  - Millions / Day
• Support different arrival rate patterns

Cyclic (Diurnal)

Growth and Spike
QueryBot 5000

SQL Workload Trace

#1 - Pre-Processor

#2 - Clusterer

#3 - Forecaster

Predicted Workload
Step #1 - Pre-Processor

• Templatization

```sql
SELECT * FROM foo WHERE id = SIGMOD
```

• Semantics equivalence check

*Millions $\rightarrow$ Thousands* 😊
**Step #2 - Clusterer**

- **Possible Similarity Features**
  - *Physical Feature*
  - *Logical Feature*
  - *Arrival Rate Feature*

![Diagram showing data flow and feature selection](image)
Coverage of the Largest Clusters

![Coverage of the Largest Clusters diagram](image)

A few large clusters exhibit major patterns
Step #3 - Forecaster

• Different models have different properties
  ▪ Linear Regression (LR), ARMA, Kernel Regression (KR), Recurrent Neural Network (RNN), FNN, PSRNN
  ▪ Properties: Linear, Memory, Kernel

• Ensemble: combine different models

LR+RNN has the best average accuracy
Prediction Results

1 Hour Horizon:

1 Week Horizon:

Bus Tracking App
Prediction Results for Spikes

ENSEMBLE: (LR+RNN)

Queries / h

Queries / h

21-Nov  1-Dec  11-Dec  21-Dec

1 Week Horizon

Admissions App

EN: 5
KR: 0

Actual  Predict

x10^6
Hybrid Model

**ENSEMBLE:** (LR+RNN)

**HYBRID:** (LR+RNN+KR)

Queries / h

Actual | Predict
--- | ---

21-Nov 1-Dec 11-Dec 21-Dec

Admissions App 1 Week Horizon

15
Example: Automatic Index Building

• Integrate QB5000 with MySQL
• Start with only primary indexes
• Same index suggestion algorithm to build 20 indexes -
  ▪ RETROSPECT: Build all indexes at once with sample history
  ▪ PREDICT: Build indexes one at a time using the forecasting
Example: Automatic Index Building

Add Index

Queries/s

6:00 9:00 12:00 15:00 18:00 21:00

Retrospect Predict

Admissions App
Takeaways

• Workload forecasting on combinations of horizons/intervals
• Reduce the forecasting cost with minimal lost of accuracy
  ▪ Templatization
  ▪ Clustering on arrival rate feature
• Hybrid forecasting method
  ▪ Capture major database workload patterns
  ▪ Maintaining good average accuracy
END

lin.ma@cs.cmu.edu

https://github.com/malin1993ml/QueryBot5000
Input Space For Kernel Regression
Related Work

• Resource Demand Prediction
• Performance Modeling and Diagnosis
• Next SQL/Transaction Prediction
• Workload Shift Detection
• Workload Compression
• Run-time Metrics Prediction (e.g. latency)