Metrics Forecasting

Dhruv Arya // Kai Franz // Neville Chima // 15-799 // Spring 2022
1. Predict future metrics based only on historical data — time series forecasting.
2. Forecasts for metrics like dead tuple percentage — can be used to intelligently schedule table vacuuming.
3. Different from MB2 — does not rely on an additional workload forecasting step.
Scope

Adds a new module metrics_forecaster

Inputs:
1. Historical metrics data (timeseries, one set per table)
2. Required granularity of predictions (e.g. 5s, 20s)
3. Required forecast length (how far into the future?)

Outputs:
1. Predicted metric values based on the historical data

Currently supported metrics: table_size, dead tuple count
Architectural Design

Control Plane

Push:
- Historical Metrics, Granularity, Desired Forecast Length, Auto-vacuum knob config

metrics_forecaster

Forecast

Neural Prophet
# Data Collected

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Benchmark Time (s)</th>
<th>Scale Factor</th>
<th>Auto-vacuum settings</th>
<th>Data Collected</th>
<th>Additional config</th>
</tr>
</thead>
<tbody>
<tr>
<td>SmallBank</td>
<td>3600</td>
<td>50/100</td>
<td>on/off</td>
<td>table size, n_dead_tuples</td>
<td></td>
</tr>
<tr>
<td>TPCC</td>
<td>3600, 7200</td>
<td>50/100</td>
<td>on/off</td>
<td>table size, n_dead_tuples</td>
<td></td>
</tr>
<tr>
<td>YCSB</td>
<td>3600, 7200</td>
<td>50/100</td>
<td>on/off</td>
<td>table size, n_dead_tuples</td>
<td>DeleteRecord weight set to 25</td>
</tr>
<tr>
<td>TATP</td>
<td>3600</td>
<td>50/100</td>
<td>on/off</td>
<td>table size, n_dead_tuples</td>
<td></td>
</tr>
</tbody>
</table>
AutoVacuum vs No AutoVacuum (YCSB)

3600 second run with a sampling rate of once per second; scale factor = 50
YCSB (Prediction 10s into the future)

**Dead tuples**

- **Autovac on**
- **Autovac off**

**Table growth**
TPC-C Table Growth Forecasting

**Autovacuum ON**

- **Table size**
  - Chart showing data, actual, and 10secondprediction.
- **Dead tuples**
  - Chart showing data, actual, and 10secondprediction.

**Autovacuum OFF**

- **Table size**
  - Chart showing data, actual, and 10secondprediction.
- **Dead tuples**
  - Chart showing data, actual, and 10secondprediction.
## Results

<table>
<thead>
<tr>
<th></th>
<th>YCSB</th>
<th>TPCC_customer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAE</td>
<td>MAE</td>
</tr>
<tr>
<td>Table Growth Vacuum ON</td>
<td>59.3</td>
<td>11.0</td>
</tr>
<tr>
<td>Table Growth Vacuum OFF</td>
<td>39.9</td>
<td>10.8</td>
</tr>
<tr>
<td>Dead Tuple Count Vacuum ON</td>
<td>79.2</td>
<td>84.9</td>
</tr>
<tr>
<td>Dead Tuple Count Vacuum OFF</td>
<td>64.4</td>
<td>49.4</td>
</tr>
</tbody>
</table>
Testing Plan

1. Store metric traces for all benchmarks in Benchbase with deletes
2. Test for training time on each benchmark:
   a. With normal trace with a 50:50 train-test split
   b. Artificially inflated trace just for test time
3. Test for correctness (MAE, RMSE):
   a. With a 50:50 split
   b. For data binned at granularity levels: 1 minute, 1 hour, 1 day, 1 week
4. Continuously log correctness metrics values per horizon evaluated
Trade-offs and Potential Problems

1. Neural prophet simply creates a linear projection for the future when predicting far into the future

2. Neural prophet takes around 13s to fit on 3600 samples of data — need to test for very large samples

3. Design does not take transferability into account. Needs to be trained from scratch on each database.
Future Work

1. Real/Longer workload traces (week, month)
2. Address the cold start problem (map to previously seen workloads?)
   1. Account for different auto-vacuum configurations
   2. Test other models: ARIMA, FBProphet, LSTMs
QUESTIONS