Cost-Aware Retraining for Machine Learning

Notations

- Current Batch $t$
- Labeled Data $D_t$
- Unlabeled Queries $Q_t$
- Model $M_t$ trained with $D_{t'}$, $t' \leq t$
- Parameters $\theta$
- Decisions:
  - Keep model $M_t$
  - Retrain model using $D_t$

Cost Matrix $C$

$$C[t', t] = \begin{cases} 
\text{Staleness Cost} & \text{if } t' < t \\
\text{Retraining Cost} & \text{if } t' = t \\
\infty & \text{otherwise}
\end{cases}$$

Staleness Cost $\overline{V}_{t,t'}$

- Cost of old model $M_t$ at batch $t$
- Query-aware performance cost
- Scenario 1: Low staleness cost
- Scenario 2: High staleness cost

Retraining Strategy $S$

- Strategy is a set of decisions
- Cost of all decisions is strategy cost
- Aim is to minimize strategy cost $S = \{\text{Keep}, \text{Keep}, \text{Retrain}, \text{Retrain}\}$

Retraining Cost $\kappa$

- Trade-off parameter
- Between resources & performance
- Low $\kappa$: performance is important
- High $\kappa$: resources are important

CARA Variants

1. CARA-T
   - $\theta$ is a threshold $\tau$
   - If staleness cost $> \tau$ then Retrain
2. CARA-CT
   - $\theta$ is a cumulative threshold $\tau_{\text{cum}}$
   - Tracks cumulative staleness cost
   - If larger than $\tau_{\text{cum}}$ then Retrain
3. CARA-P
   - $\theta$ is a periodicity $\phi$
   - Retrain decision after every $\phi$ batches

Results

<table>
<thead>
<tr>
<th>$\theta^*$ of Cara variants for CovCon-D</th>
<th>$\tau$</th>
<th>$\tau_{\text{cum}}$</th>
<th>$\phi^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>如意实用</td>
<td>$1170$</td>
<td>$11.20$</td>
<td></td>
</tr>
<tr>
<td>如意实用-H</td>
<td>$1507$</td>
<td>$13.89$</td>
<td></td>
</tr>
<tr>
<td>如意实用-CT</td>
<td>$1368$</td>
<td>$13.33$</td>
<td></td>
</tr>
<tr>
<td>如意实用-P</td>
<td>$1248$</td>
<td>$10.75$</td>
<td></td>
</tr>
</tbody>
</table>

Ananth Mahadevan$^1$  Michael Mathioudakis$^1$

$^1$ Department of Computer Science, University of Helsinki