Statistical Model Updates in Distributed Computing: An Optimal Stopping Theory Perspective

Ekaterina (Katie) Aleksandrova
2133352A@student.gla.ac.uk

FRI 12th APRIL @ SoCS/SAWB 303  Supervision by: Dr Chris Anagnostopoulos
Problem statement

- **Internet of Things system:**
  - Edge sensors
  - Neighbourhood edge gateways
  - Data centres (the Cloud)

- **What we Are doing:**
  - Sense multivariate contextual data at the edge
  - Transfer the data to the Cloud for analysis
  - Have accurate and up-to-date knowledge in the Cloud

- **What we Don’t want:**
  - Computational overhead at the Data centres and sensors
  - Communication overhead
  - High network bandwidth
Problem statement (Cont’d)

- **What we Can do:**
  - Gather some of the sensed data in the sensor
  - Create a model from that data
  - Communicate the model
  - Wait until a Model Concept Drift (CD) has occurred
  - Communicate an updated model

- **What we Will achieve:**
  - Less communication in the network
  - Lower bandwidth requirement
  - Data is delivered to the datacentre partially analysed
  - Data is anonymised by preserving the raw context at the sensor level
What is a Concept Drift?

**Def.** A changing context which induces a change in the target concepts (Widmer & Kubat, 1996)

(Lemaire et. al., 2015)
Handling concept drift
Using Cumulative Sum (CuSum)\([\text{[*]}]\)

- Using the Absolute Error Difference between current model and previously sent model on the most up-to-date subset of the data:
  \[ \Delta e = |e - e'| \]

- Make assumption on the **Good Distribution** and the **Bad distribution**

- Calculate the probability density functions of both:
  - \( P_{\text{good}} \) and \( P_{\text{bad}} \)

\([\text{[*]}]\) Invented by E. Page, Uni of Cambridge, 1954

Fig. **Good Distribution vs. Bad Distribution**
Handling concept drift
Using Cumulative Sum (CuSum) (Cont’d)

- For each new $\Delta e$ calculate the log-likelihood ratio:

  $$l_t = L_{\Delta e} = \ln \frac{P_{\Delta e | \text{bad}}}{P_{\Delta e | \text{good}}}$$

- Keep a record of all log-likelihood ratios sums:

  $$S[t] = \sum_{k=0}^{t} l_k$$

- Decision value:

  $$g = S[t] - \min_{0 \leq k \leq t-1} (S[k])$$

- Update criteria: $g > h$, where $h \leftarrow \text{threshold}$
From CuSum to Optimal Stopping Theory

- What does Optimal Stopping Theory deal with?
  - How to estimate the best time to stop and gain the highest reward or suffer the least penalty?

- Popular examples:
  - The Secretary problem
  - The Blackjack Card game
  - The House Selling problem

- Our problem:
  - Delay sending an update as much as possible until a change in the distribution has occurred
From CuSum to OST (Cont’d)

- We use the cumulative sum principle on the absolute error rate, which is not allowed to exceed $\Theta$
  - $Z_t = \Delta e_t = |e_t - e'_t|
  - $S_t = \sum_{k=0}^t Z_k$

- Optimise $t^*$ while maximising the reward function
  $$V_t = \begin{cases} t, S_t \leq \Theta; & \text{if we didn't update and the next } \Delta e \text{ was low} \\ -B, S_t > \Theta; & \text{if we didn't update and the next } \Delta e \text{ was high} \end{cases}$$

- From here we can obtain the expected reward at time $t$
  $$\Rightarrow \mathbb{E}[V_t] = t \cdot P(S_t \leq \Theta) + (-B \cdot P(S_t > \Theta))$$
  $$= (t + B) \cdot P(S_t \leq \Theta) - B$$

- Given the realisation of all random variables up to time $t$, let us have the filtration
  $$\mathbb{F}_t = \{S_1, ..., S_t\} \cup \{Z_1, ..., Z_t\}$$

- From that we can express the conditional expectation of the future reward $V_{t+1}$
  $$\mathbb{E}[V_{t+1}|\mathbb{F}_t] = (t + B) \cdot P(S_{t+1} \leq \Theta|\mathbb{F}_t) - B$$
The equation $S_t = \sum_{k=0}^{t} Z_k$ can be used to express the sum at time $t + 1$:

$$S_{t+1} = \sum_{k=0}^{t+1} Z_k = \sum_{k=0}^{t} Z_k + Z_{t+1} = S_t + Z_{t+1}$$

This leads to finding that the **probability of the future sum** being less than $\theta$ given our filtration equals the **cumulative distribution function** of $Z$ being less than or equal to $\theta - S_t$:

$$P(S_{t+1} \leq \theta | \mathcal{F}_t) = P(S_t + Z_{t+1} \leq \theta | \mathcal{F}_t)$$
$$= P(Z_{t+1} \leq \theta - S_t | \mathcal{F}_t)$$
$$= F_z(\theta - S_t)$$

Now we can substitute for the conditional expectation of the future value and obtain:

$$\mathbb{E}[V_{t+1} | \mathcal{F}_t] = (t + 1 + B) \cdot F_z(\theta - S_t) - B$$

When the **currently obtained reward** is **more** than the **conditional expected future reward**, we want to send an updated model. That is when the following is satisfied:

$$V_t \geq \mathbb{E}[V_{t+1} | \mathcal{F}_t] \iff F_z(\theta - S_t) \leq \frac{t + B}{t + 1 + B}$$
Other update policies and their rationale

- **Median-based policy**
  - **Update criteria:** $\Delta e_t > \alpha \times \text{median} (\Delta e_1, ..., \Delta e_{t-1})$

- **Accuracy-based policy**
  - **Update criteria:** $e_t > e'_t$

- **Random-based policy**
  - **Update criteria:** $\text{generate}\{1|0\}, \text{prob} = \text{optimal policy}$
Experimenting using real life sensor data

- **GNFUV Unmanned Surface Vehicles Sensor Dataset**
  (Harth & Anagnostopoulos, 2018)
  - **collected data**: (humidity, temperature)
  - used with **Linear Regression** models

- **Gas Sensors for Home Activity Monitoring Dataset**
  (Huerta et. al., 2016)
  - **collected data**: (humidity, temperature, values from 8 metal-oxide sensors)
  - used with **Support Vector Regression** (with an RBF kernel) models
  - included artificial incremental concept drift in the data
Our results for Linear Regression Models

- The **absolute error** for the optimal policy does not drastically deviate from the other policies.

- The optimal policy saves on average **5 times** more communication.
ANOVA test for Linear Regression

Looking for a statistical significant difference b/w policies:

waiting time

<table>
<thead>
<tr>
<th>ANOVA p-value for waiting time</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sensor pi3</td>
<td>1.248e-30</td>
<td>&lt;= 0.05</td>
</tr>
<tr>
<td>sensor pi4</td>
<td>7.893e-14</td>
<td>&lt;= 0.05</td>
</tr>
</tbody>
</table>

absolute error

<table>
<thead>
<tr>
<th>ANOVA p-value for abs error</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sensor pi3</td>
<td>1.244e-13</td>
<td>&lt;= 0.05</td>
</tr>
<tr>
<td>sensor pi4</td>
<td>2.723e-17</td>
<td>&lt;= 0.05</td>
</tr>
</tbody>
</table>
Tukey’s HSD test for Linear Regression

The waiting time for the optimal policy has a higher mean and the difference is statistically significant.

The difference in the absolute error between the optimal policy and when always having the most up-to-date model is not statistically significant.
Our results for Support Vector Regression Models

- The absolute error for the optimal policy and the cusum policy deviate the most from the accurate policy.

- The optimal policy waits on average 30 times longer than the other policies.
ANOVA test for Support Vector Regression Models

Looking for a statistical significant difference b/w policies:

- waiting time

<table>
<thead>
<tr>
<th>sensor</th>
<th>p-value</th>
<th>p-value ≤ 0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>R3</td>
<td>9.198e-18</td>
<td>≤ 0.05</td>
</tr>
<tr>
<td>R5</td>
<td>1.477e-32</td>
<td>≤ 0.05</td>
</tr>
</tbody>
</table>

- absolute error

<table>
<thead>
<tr>
<th>sensor</th>
<th>p-value</th>
<th>p-value ≤ 0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>R3</td>
<td>6.003e-23</td>
<td>≤ 0.05</td>
</tr>
<tr>
<td>R5</td>
<td>1.062e-11</td>
<td>≤ 0.05</td>
</tr>
</tbody>
</table>
Tukey’s HSD test for Support Vector Regression Models

The waiting time for the optimal policy has a higher mean and the difference is statistically significant.

The optimal policy has a higher absolute error rate and the difference is statistically significant from the rest of the policies.
## Conclusion

<table>
<thead>
<tr>
<th>Policy type</th>
<th>High quality prediction models</th>
<th>Lower quality prediction model</th>
</tr>
</thead>
<tbody>
<tr>
<td>CuSum</td>
<td>high error and high communication 😞</td>
<td></td>
</tr>
<tr>
<td>Accuracy-based</td>
<td>😊</td>
<td>high communication</td>
</tr>
<tr>
<td>Optimal</td>
<td>high error</td>
<td>😞</td>
</tr>
<tr>
<td>Median-base</td>
<td>😊</td>
<td>high communication</td>
</tr>
</tbody>
</table>
Thank you!

Ekaterina (Katie) Aleksandrova