University of Glasgow

School *of* Computing Science
Knowledge & Data Engineering Systems

Maintenance of Model Resilience in Distributed Edge Learning Environments

<u>Qiyuan Wang</u>, Christos Anagnostopoulos, Jordi Mateo Fornes, Kostas Kolomvatsos, Andreas Vrachimis

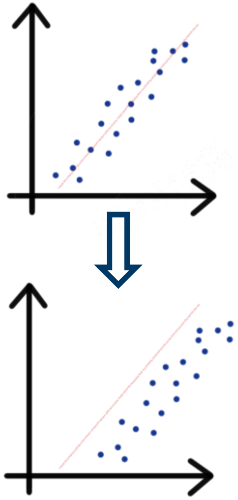
WORLD CHANGING GLASGOW

27-30 JUNE 2023 @ 19th International Conference on Intelligent Environments (IE’23)

Quick Review

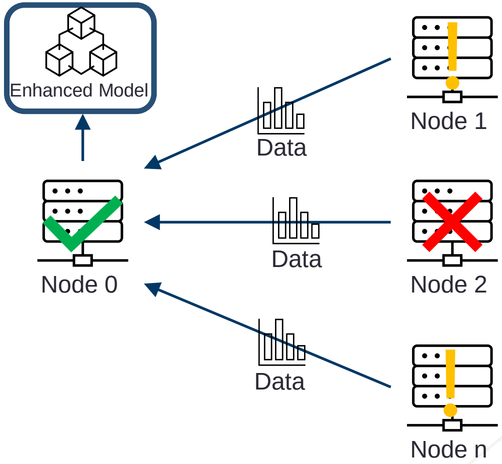
Problem

Concept Drift



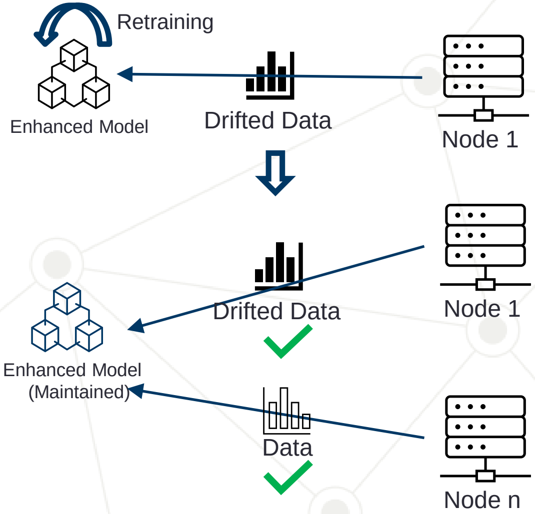
Environment

“Enhanced” Models



Target

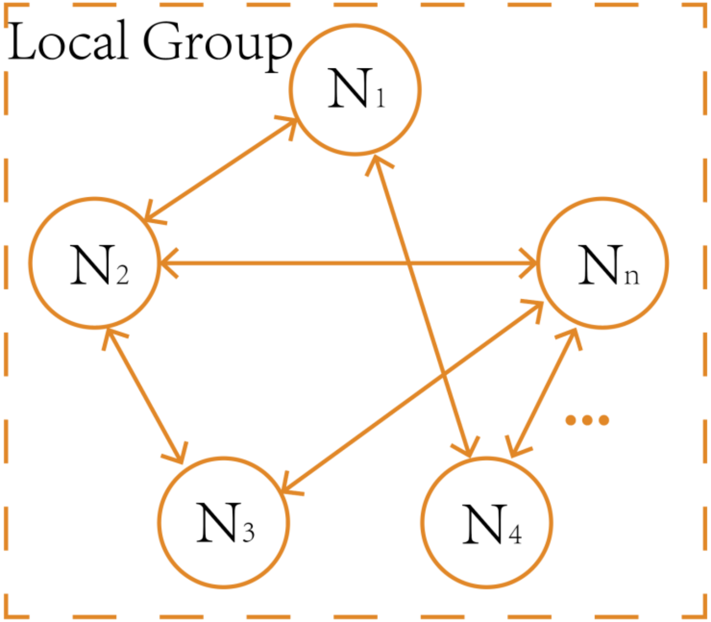
Successful Maintenance



Background

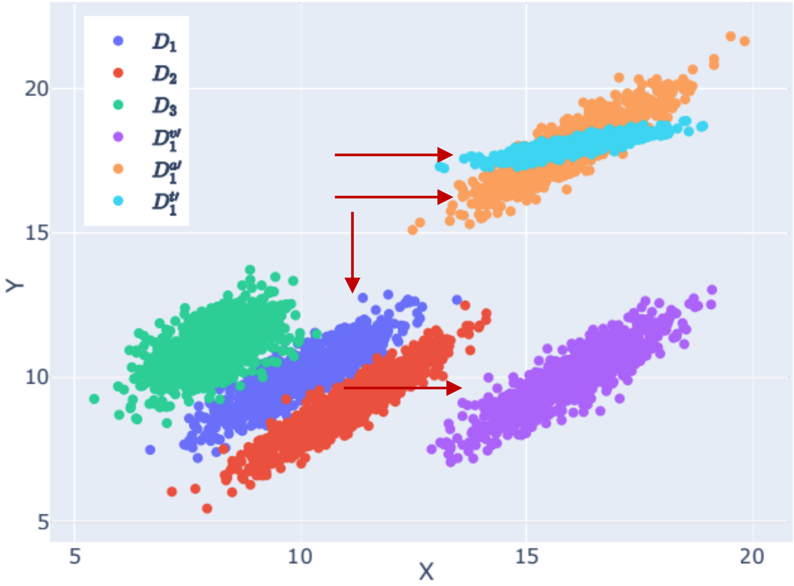
System Formulation

- EC system with n distributed nodes: $\mathcal{N} = \{N_1, \ldots, N_n\}$
- Node N_i has its own local data $D_i = \{(x, y)_l\}_{l=1}^{L_i}$, with L_i input-output pairs $(x, y) \in \mathcal{X} \times \mathcal{Y}$
- The input $x = [x_1, \ldots, x_d]^\top \in R^d$ is a d-dim feature vector, which is assigned to output $y \in \mathcal{Y}$ used for regression (e.g., $\mathcal{Y} \subseteq R$) or classification predictive tasks (e.g., $\mathcal{Y} \subseteq \{-1,1\}$)
- The neighbourhood of N_i: $\mathcal{N}_i \subseteq \mathcal{N} \setminus \{N_i\}$



Background

Predictive Services: Regression



Drift Classification

Virtual Drift:

$$P(x) \neq P(x') \land P(y) = P(y')$$

Actual Drift:

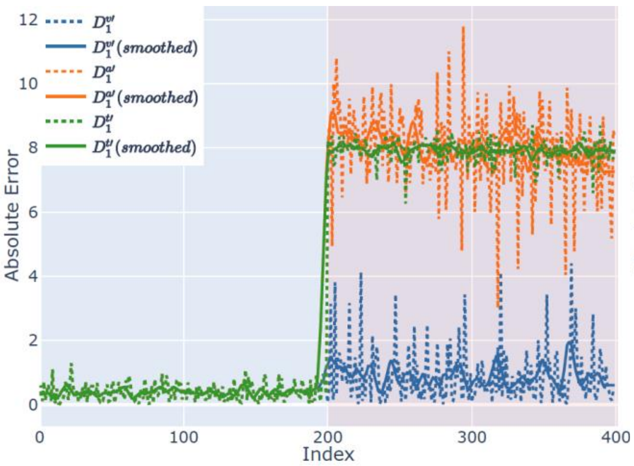
$$P(x) \neq P(x') \land P(y) \neq P(y')$$

Total Drift:

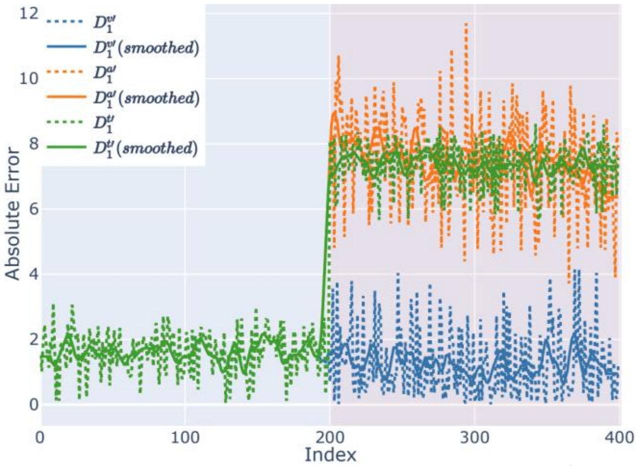
$$P(x) \neq P(x') \land P(y) \neq P(y') \land P(y \mid x) \neq P(y' \mid x')$$

School *of* Computing Science
Knowledge & Data Engineering Systems

Effects of Concept Drifts



Performance of local model f_1



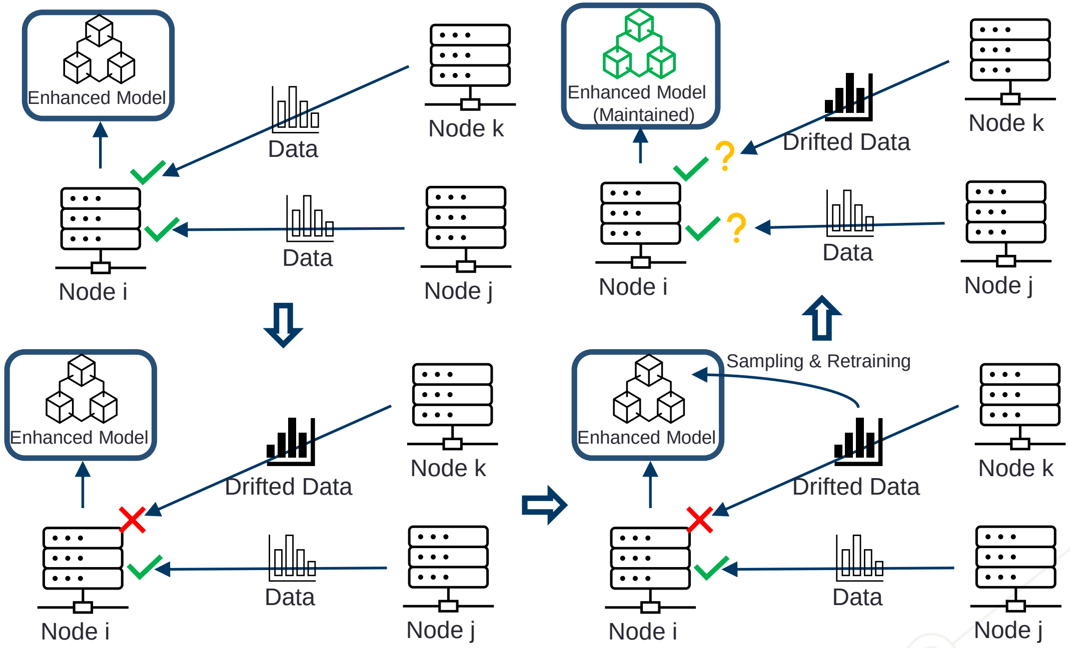
Performance of enhanced model $\bar{f}_2$ (SVR)

TABLE I
PERFORMANCE OF DIFFERENT MODELS

	RMSE			
Model	D_1	$D_1^{v\prime}$	$D_1^{a\prime}$	$D_1^{t\prime}$
f_1	0.47	1.29	8.06	7.93
$\bar{f}_2^{GS}(SVR)$	1.73	1.69	7.57	7.41
$\bar{f}_2^{CG}(SVR)$	1.69	1.68	7.57	7.41
$\bar{f}_2^{GS}(GBR)$	1.54	2.19	6.26	6.12
$\bar{f}_2^{CG}(GBR)$	1.50	2.17	6.29	6.15

$D_1^{v\prime}$ corresponds to virtual drifted D_1
$D_1^{a\prime}$ corresponds to actual drifted D_1
$D_1^{t\prime}$ corresponds to total drifted D_1

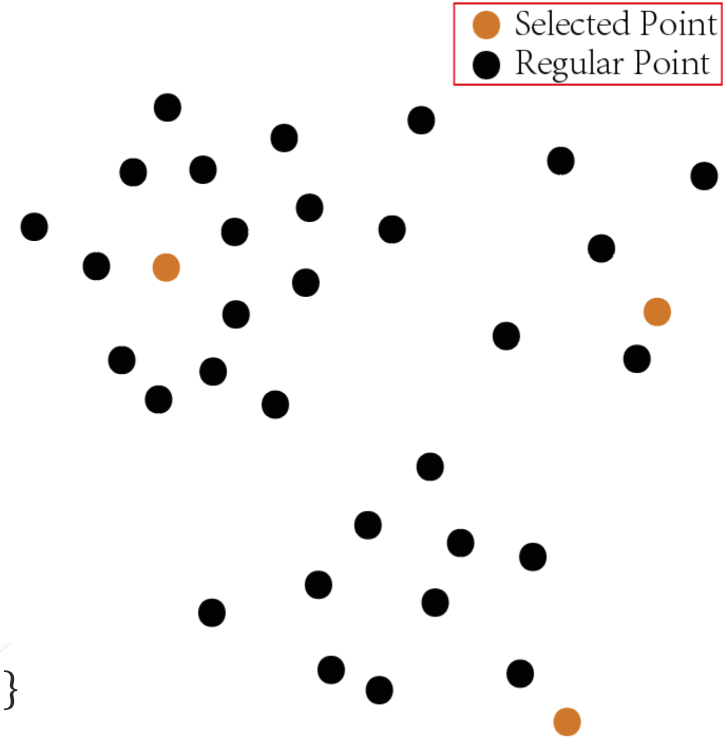
Objective



- O1: Minimize $E\mathcal{L}\left(\bar{f}_i'(D_k{}')\right)$ (for node N_k)
- O2: Minimize $E\mathcal{L}\left(\bar{f}_i'(D_j)\right)$ (for node N_j)
- O3: Reduce inter-node data transfer between nodes N_i and N_k (during maintenance)

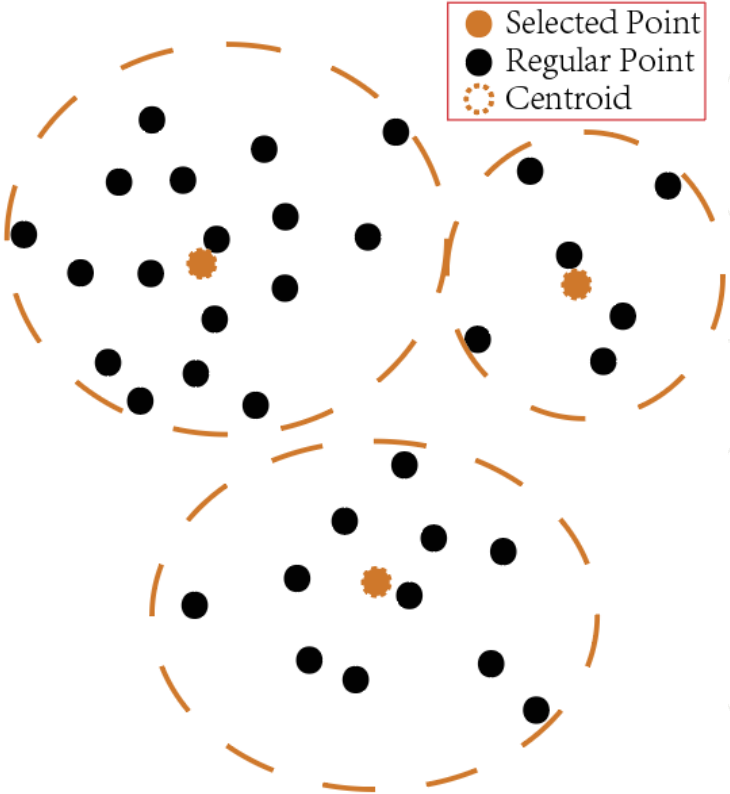
Model Maintainability Strategies – GS (**Sampling**)

- N_i: Node with **enhanced** model, with local data D_i
- N_j: Node to be **surrogated**, with local data D_j
- In this context, $N_j = N_k$, $D_j = D_k{}'$
- Based on **random sampling** of D_j, i.e., $\Gamma\left(D_j\right) \subset D_j$
- Sample mixing rate $\alpha = \frac{|\Gamma(D_j)|}{|D_j|} \in (0,1)$, controlled by N_i
- Incremental learning supported?
  - **Yes**: maintain model with $\Gamma\left(D_j\right)$
  - **No**: Training from **scratch** with $\overline{D_i}' = \overline{D_i} \cup \{\Gamma\left(D_j\right)\}$



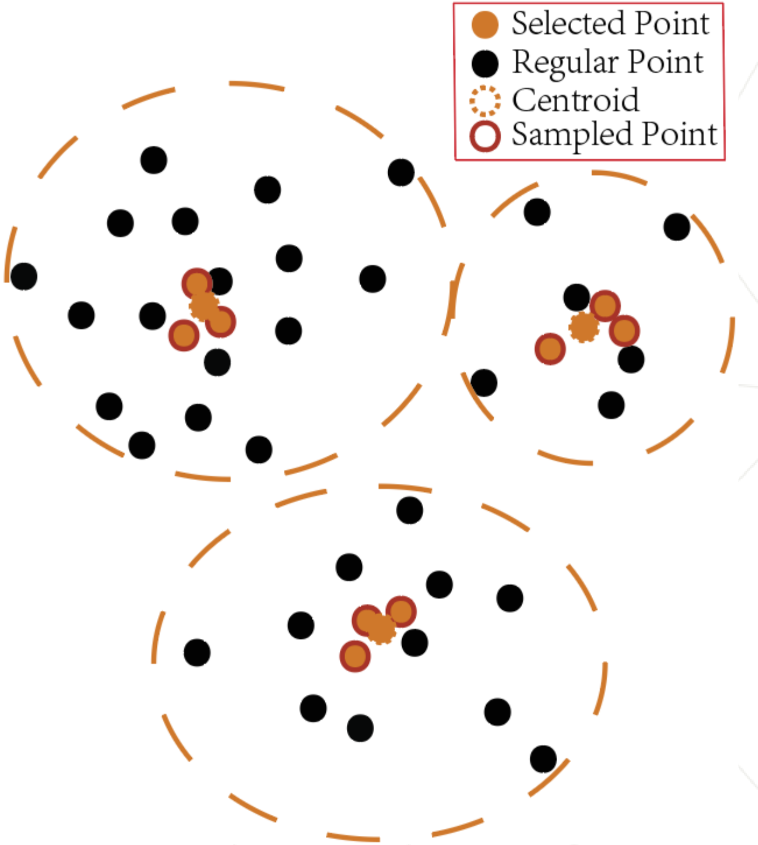
Model Maintainability Strategies – CG (**Centroid**)

- Input-output space $\mathcal{X} \times \mathcal{Y}$ of D_j is **quantized**
- The number of clusters K depends on the size $L_i = |D_j|$ and **mixing rate** $\boldsymbol{\alpha}$ i.e., $K = \alpha|D_j|$
- $\Gamma(D_j) = \cup_{k=1}^{K} \{w_{jk}\}$
- Does <u>not</u> transfer real data



Model Maintainability Strategies – ECG (**Centroid+**)

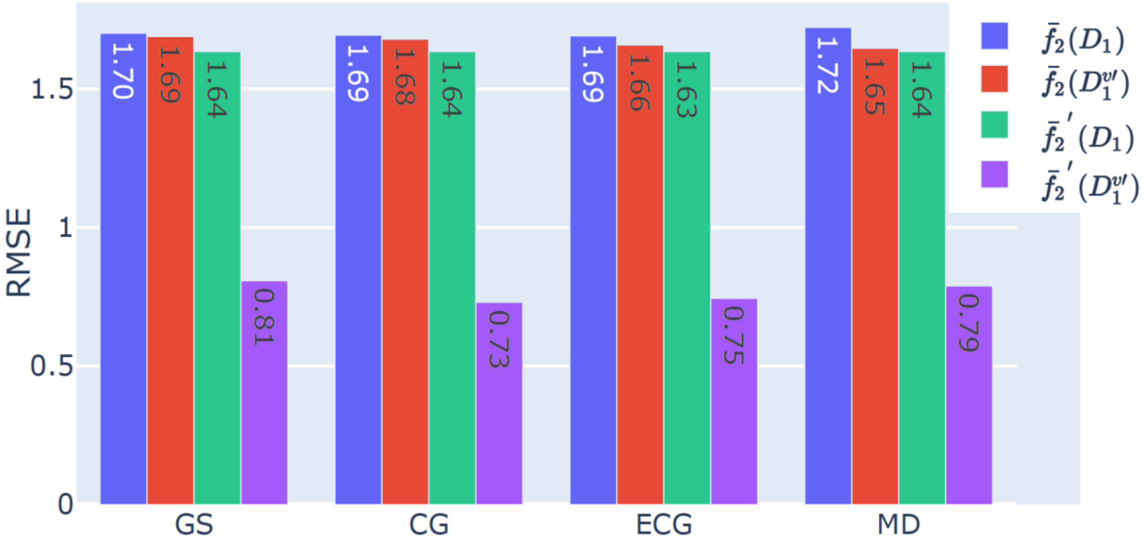
- Input-output space $\mathcal{X} \times \mathcal{Y}$ of D_j is **quantized**
- λ introduced to control the **duplication** of centroids
- The number of clusters $K = \frac{\alpha|D_j|}{\lambda}$
- For each cluster, sample the **centroid and $\lambda - 1$ points** $(\hat{x}, \hat{y})$ from $\mathcal{N}\left(w_{jk}, \sigma_j^2\right)$
- $\Gamma\left(D_j\right) = \cup_{k=1}^{K} \{w_{jk} \cup \{(\hat{x}, \hat{y}) \sim \mathcal{N}\left(w_{jk}, \sigma_j^2\right)\}\}$
- Does <u>not</u> transfer real data



Model Maintainability Strategies – MD (**Generative Data**)

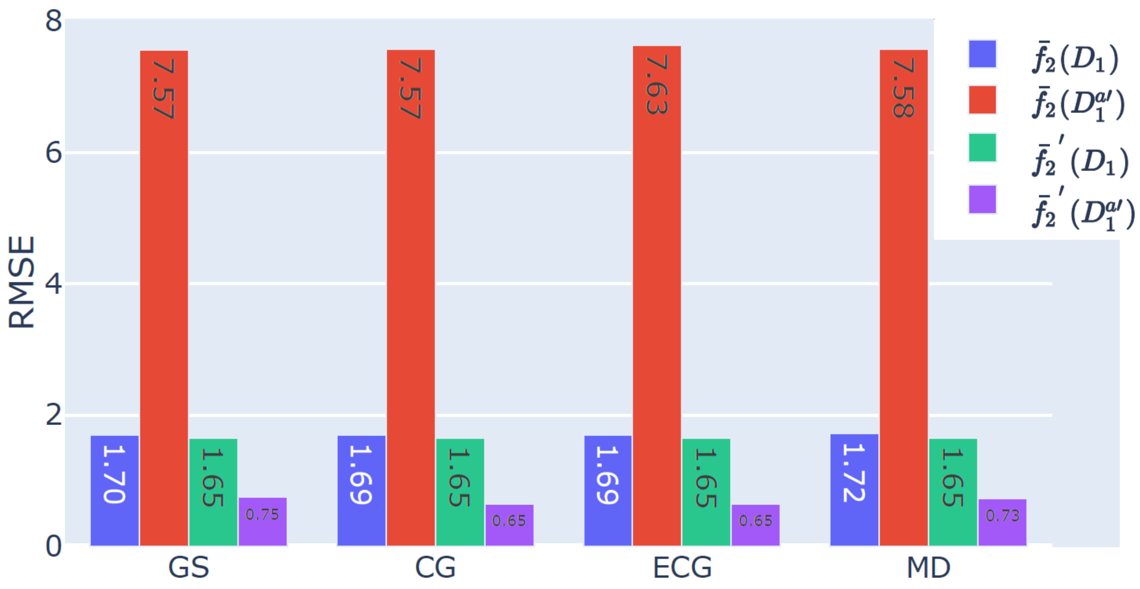
- $\mu_j = \frac{\sum_{m=1}^{|D_j|} x_m}{|D_j|} \in R^d$, $\sigma_j = \sqrt{\frac{1}{|D_j|}\sum_{m=1}^{|D_j|}(x_m - \mu_j)^2} \in R^d$, **SEM** $\bar{\sigma}_j = \frac{\sqrt{\frac{1}{|D_j|}\sum_{m=1}^{|D_j|}\left(y_m - \frac{\sum_{m=1}^{|D_j|} y_m}{|D_j|}\right)^2}}{\sqrt{|D_j|}} \in R^d$
- $\boldsymbol{\mu_j}$, $\boldsymbol{\sigma_j}$ and $\boldsymbol{\bar{\sigma}_j}$, alongside with $\mathbf{f_j}$ are sent to N_i for mock data generation
- ϵ_j: random noise sampled from $\boldsymbol{\mathcal{N}(0, \bar{\sigma}_j^2)}$
- $\Gamma(D_j) = \{(\widehat{\mathcal{X}_j}, \widehat{\mathcal{Y}_j}) : \widehat{\mathcal{X}_j} \sim \mathcal{N}(\mu_j, \sigma_j^2), \widehat{\mathcal{Y}_j} = f_j(\widehat{\mathcal{X}_j}) + \epsilon_j\}$
- Does <u>not</u> transfer real data

Experiments & Evaluation – Virtual Drifts



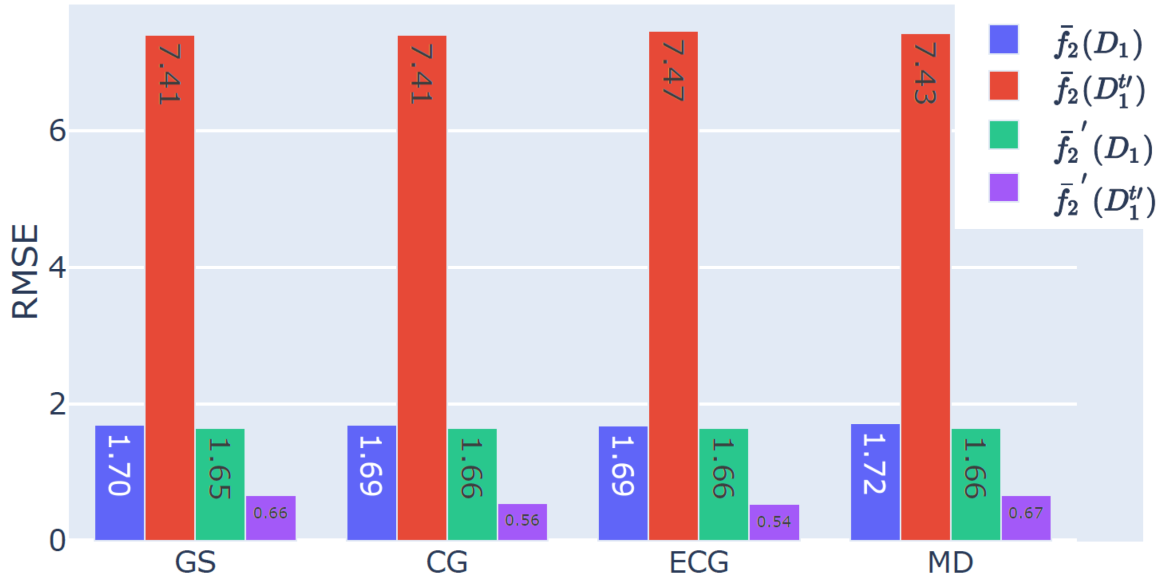
- Virtual drift did not affect the performance of $\bar{f}_2$ negatively (red bar vs blue bar)
- Maintenance was able to improve the performance on $D_1^{v\prime}$ further (purple bar vs red bar) while keeping the performance on D_1 (green bar vs blue bar)
- **CG & ECG** are the **best** strategies overall

Experiments & Evaluation – Actual Drifts



- $\bar{f}_2$ could not handle the actual drift without maintenance (red bar vs blue bar)
- Maintenance was very effective, drastically improved the performance on $D_1^{a'}$ (purple bar vs red bar) while keeping the performance on D_1 (green bar vs blue bar)
- **CG & ECG** are the **best** strategies overall

Experiments & Evaluation – Total Drifts



- $\bar{f}_2$ could not handle the total drift without maintenance (red bar vs blue bar)
- Maintenance was very effective, drastically improved the performance on $D_1^{t'}$ (purple bar vs red bar) while keeping the performance on D_1 (green bar vs blue bar)
- **ECG** is the **best** strategy overall

Experiments & Evaluation – Effects on Other Node(s)

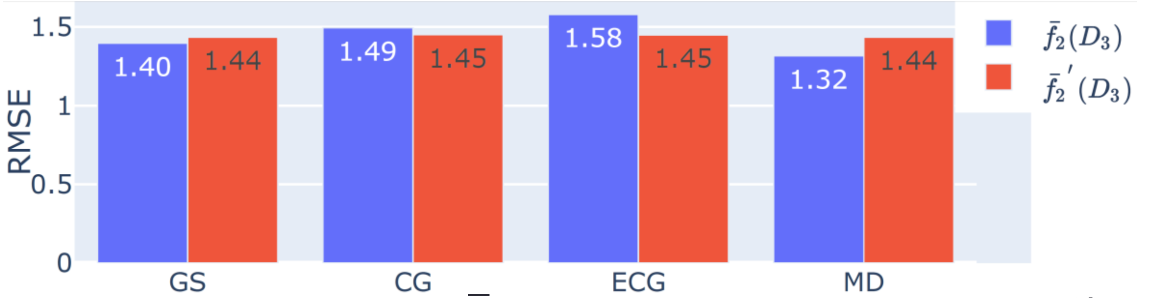
Performance (on D_1& D_3) of the initial enhanced model $\overline{f}_2$

Performance (on D_3) of $\overline{f}_2$ before & after maintenance with $D_1^{v\prime}$

- Strategies used to build the enhanced model initially affect the performance of $\overline{f}_2$ on D_3

- D_3 is almost **indifferent** to the strategies used for the maintenance

Experiments & Evaluation – Effects on Other Node(s)

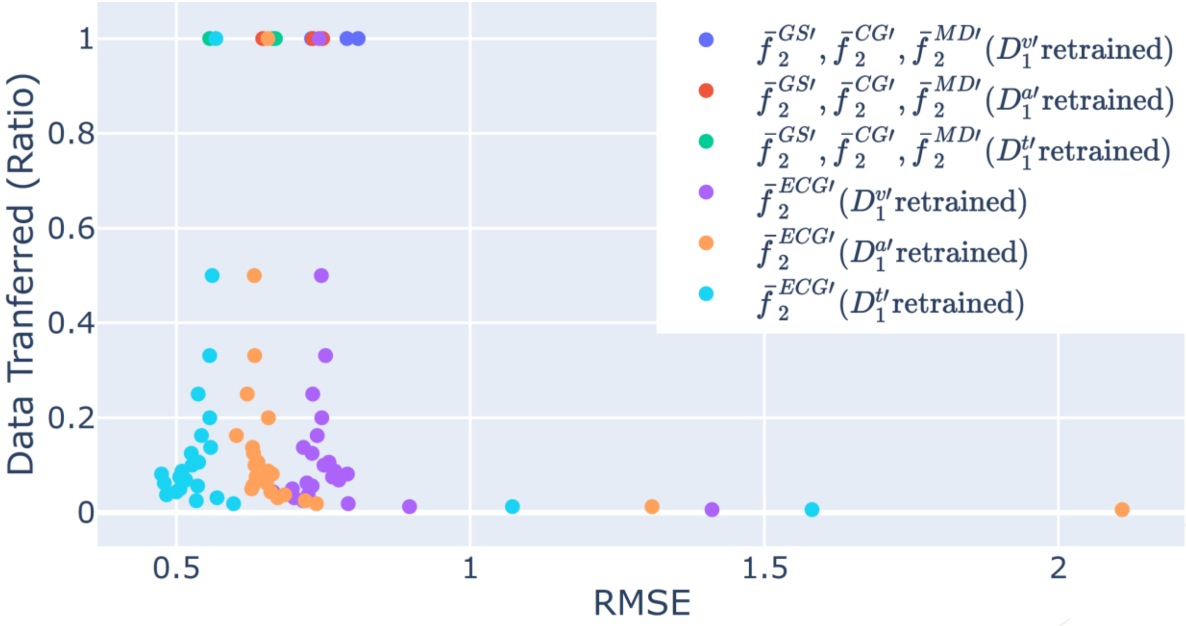


Performance (on D_1) of $\overline{f_2}$ before & after maintenance with $D_1^{a\prime}$

Performance (on D_1) of $\overline{f_2}$ before & after maintenance with $D_1^{t\prime}$

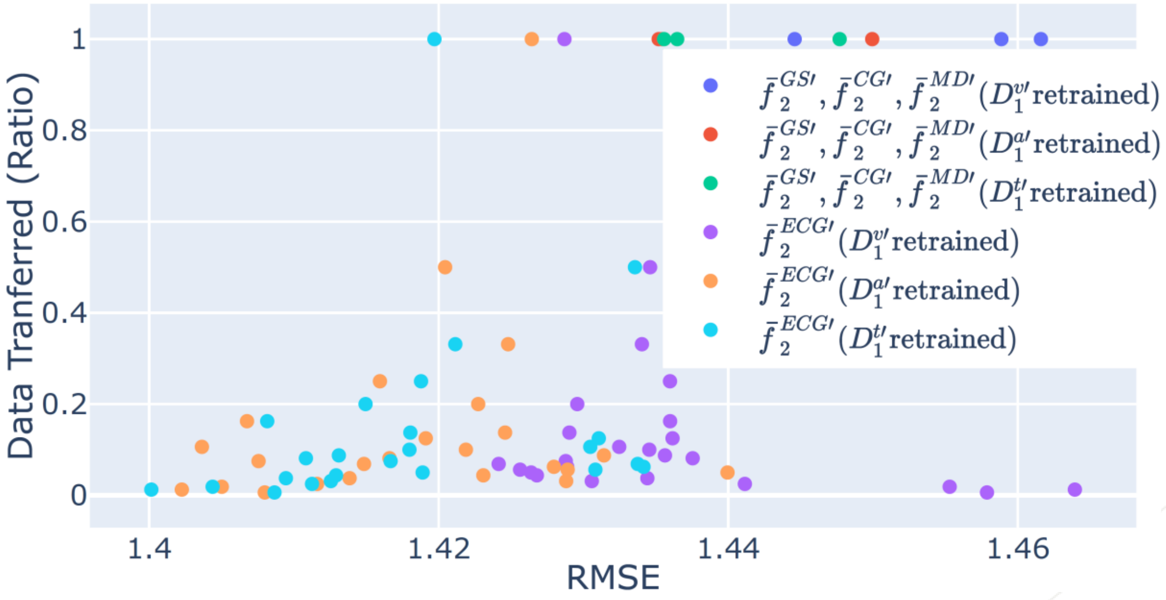
- Identical results for maintenance with $D_1^{a\prime}$ and $D_1^{t\prime}$
- For all 3 kind of drifts, the maintenance did not affect the other node

Experiments & Evaluation – Data Transfer



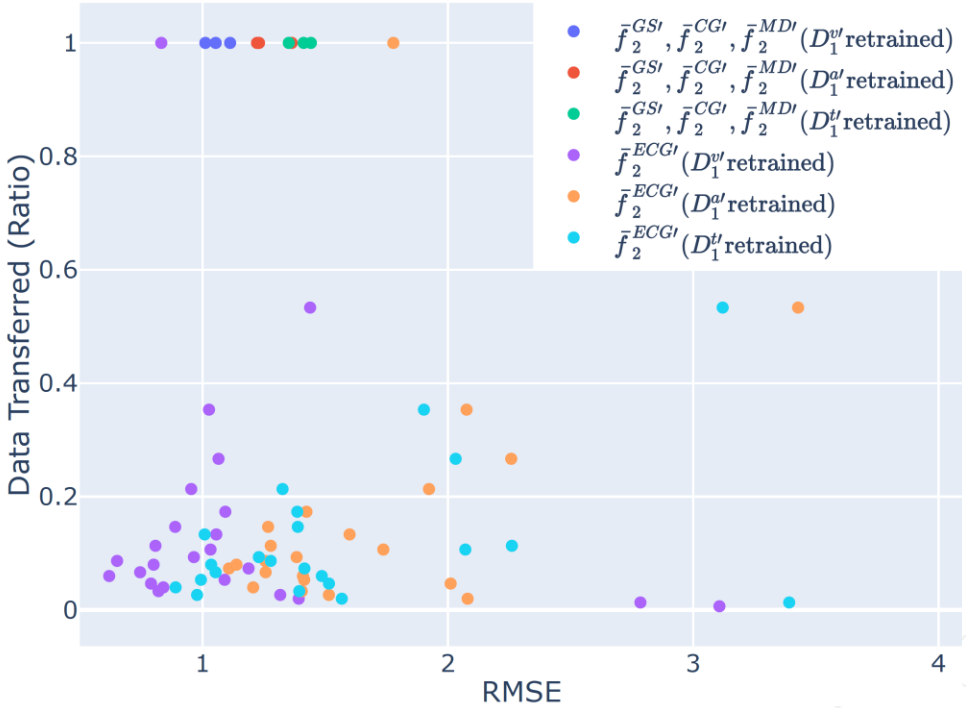
- Performance on $D_1^{v\prime}$, $D_1^{a\prime}$, $D_1^{t\prime}$
- Given the same α, both GS and CG transfer the same amount of data
- For MD, the statistics and the model need to be transferred are at the **same scale** of GS and CG
- For ECG, we manipulate ***intensity*** $\boldsymbol{\lambda}$ to directly control the amount of transferred data
- Ideal: bottom left

Experiments & Evaluation – Data Transfer



- Performance on D_3
- The magnitude of variation in performance is **negligible**
- ECG is working very well in both reducing the data transferred and maintaining the performance

Experiments & Evaluation – Data Transfer



- Results got with **realistic** dataset: $GNFUV^*$
- Similar results to what we got before
- Only 5% - 10% data transfer needed for ECG to achieve the best performance

*: https://archive.ics.uci.edu/dataset/452/gnfuv+unmanned+surface+vehicles+sensor+data

Conclusions

- Investigated the problem of maintaining resilient **enhanced models** in **DML** environments
- Proposed 4 model maintainability **strategies**
- Evaluated the effects of these strategies on 3 kinds of drifts
- Proved the **effectiveness** and **efficiency** of proposed approaches

University of Glasgow

School *of* Computing Science
Knowledge & Data Engineering Systems

Thank you!

Qiyuan Wang
Qiyuan.Wang@glasgow.ac.uk