**Supplement Analysis of PMH-DDM**

In the manuscript, the evidence for the proposed Parallel Multi-hypothesis Drift Detection (PMH-DDM) is shown in the ablation study section (Section 5.5 in the modified manuscript). The ablation study (D) shows the performance of CSAL when the target stream drift detection mechanism is absent. Results in Table 7 shows that illustrate a 0.3% decrease in the weighted labeling cost with the absence of PMH-DDM, but a 2.8% decrease in the overall average accuracy.

We provide more detailed experimental results here for reference:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | CSAL | | | | AS(D) | | | |
|  | acc(%) | lt(%) | ls(%) | lw(%) | acc (%) | lt(%) | ls(%) | lw(%) |
| T1S | **97.1** | 11.8 | 20.7 | 18.7 | 92.6 | 16.3 | 20.8 | 23.3 |
| T1G | 95.1 | 23.2 | 19.6 | 29.7 | **95.3** | 9.1 | 19.3 | 15.5 |
| T1I | **97.6** | 12.2 | 21.0 | 19.2 | 96.9 | 8.0 | 20.8 | 14.9 |
| T2S | **91.0** | 12.0 | 12.3 | 16.1 | 90.7 | 25.7 | 12.0 | 29.7 |
| T2G | **91.5** | 12.2 | 12.0 | 16.2 | 89.2 | 5.5 | 12.0 | 9.5 |
| T2I | 93.8 | 15.7 | 12.1 | 19.8 | **94.9** | 12.5 | 12.1 | 16.5 |
| T3S | **95.4** | 12.0 | 11.8 | 15.9 | 91.3 | 40.3 | 11.8 | 44.3 |
| T3G | **92.5** | 30.6 | 12.3 | 34.7 | 89.0 | 17.6 | 12.3 | 21.7 |
| T3I | **93.9** | 16.2 | 12.1 | 20.2 | 93.6 | 29.1 | 12.2 | 33.2 |
| Ws3 | **71.3** | 15.6 | 20.4 | 22.4 | 70.5 | 16.0 | 19.6 | 22.5 |
| Ws6 | 70.1 | 14.1 | 23.2 | 21.8 | **70.3** | 11.9 | 27.0 | 20.9 |
| Ws9 | 69.7 | 14.2 | 22.4 | 21.6 | **69.9** | 14.5 | 24.0 | 22.5 |
| Wns3 | **71.7** | 16.8 | 21.5 | 24.0 | 68.5 | 10.4 | 24.6 | 18.6 |
| Wns6 | **71.6** | 17.9 | 25.6 | 26.4 | 69.2 | 11.6 | 24.3 | 19.7 |
| Wns9 | **70.8** | 20.0 | 22.9 | 27.7 | 70.0 | 3.1 | 22.9 | 10.7 |
| holiday | 79.8 | 25.6 | 21.1 | 32.6 | **80.2** | 39.5 | 20.9 | 46.5 |
| weekend | **76.4** | 21.9 | 15.7 | 27.2 | 62.8 | 12.1 | 15.8 | 17.4 |
| weekday | **70.2** | 19.3 | 15.9 | 24.6 | 55.7 | 31.6 | 15.8 | 36.9 |
| Industry | **95.3** | 14.4 | 12.7 | 18.6 | 94.6 | 3.7 | 13.1 | 8.1 |
| Average | **83.9** | 17.1 | 17.6 | 23.0 | 81.3 | 16.8 | 18.0 | 22.8 |

From the table, it can be concluded that PMH-DDM can improve the accuracy scores of CSAL both on the synthetic and real datasets.

Furthermore, we provide here a specific analysis of the drift detection performance of the PMH-DDM based on the synthetic datasets used in experimental comparison. According to some relevant literatures [1-4], we chose the following metrics to evaluate the concept drift detection.

**False Positive Rate (FP)**: claimed changes when concept drift does not really happen.

**False Negatives Rate (FN)**: changes missed when concept drift does occur.

**Delay**: distance between the concept drift detection point and the actual drift point.

The descriptions of the datasets are shown in Section 5.1 of the manuscript. Since gradual and incremental concept drift datasets are included in comparison, we set some basis rules in our experimental analysis. For the sudden concept drift dataset, the most recent drift report after the change point is considered to be a correct report, and all other drift reports are treated as false positives. On the gradual concept drift and incremental concept drift datasets, concept drifts report outside the drift occurrence range are treated as false alarms. The delay indicator is derived by calculating the distance between the drift start position and the closest drift reporting position after the start position.

Therefore, First, the detection of mutational concept drift only identifies the closest drift detection report to the drift point after the occurrence of concept drift, and the others are treated as false positives.

On both the asymptotic concept drift and incremental concept drift datasets, concept drifts outside the concept drift occurrence interval are treated as false positives. The distance of conceptual drift was derived by calculating the location where the drift first occurred and the most recent drift report.

PMH-DDM is mainly used to deal with the drift detection problem in active learning condition. Therefore, results of (labeling cost = 0.3) and are included in comparison. The selection of the base models and the parameter settings are consistent with the experimental part of the manuscript. The experimental results were obtained from ten parallel experiments.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | CSAL with PMH-DDM | | |  | | |  | | |
|  | **FP** | **FN** | **delay** | **FP** | **FN** | **delay** | **FP** | **FN** | **delay** |
| T1S | 0.29 | 0 | 97 | 0 | 0 | 75 | 0 | 0 | 113 |
| T1G | 0.25 | 0 | 134 | 0 | 0 | 366 | 0.18 | 0 | 179 |
| T1I | 0.05 | 0 | 1933 | 0 | 1 | - | 0 | 1 | - |
| T2S | 0.09 | 0 | 966 | - | 0 | 123 | 0 | 0 | 9 |
| T2G | 0.16 | 0 | 1769 | - | 0 | 454 | 0 | 0 | 193 |
| T2I | 0 | 0 | 1360 | - | 0 | 5618 | 0 | 0 | 4804 |
| T3S | 0.09 | 0 | 118 | - | 0 | 97 | 0 | 0 | 22 |
| T3G | 0 | 0 | 135 | 0.1 | 0 | 257 | 0 | 0 | 399 |
| T3I | 0.02 | 0 | 913 | - | 1 | - | - | - | - |

As shown in the table, PMH-DDM has a relatively good conceptual drift detection capability. Particularly, CSAL with PMH-DDM shows better drift detection ability on incremental concept drift. On T1I and T3I, the and fail to detect the concept drift. And the average delay of PMH-DDM is smaller than and on T2I.

A possible weakness of PMH-DDM is the false alarm rate, since PMH-DDM uses a confidence-based approach for concept drift detection. Although this approach is beneficial in capturing the changes in model confidence due to concept drift, compared to error rate-based approaches that are prone to possible false positives [5]. Another possible weakness of PMH-DDM is the concept drift detection capability of decision boundary drift (T2 datasets), which may be due to the small amount of target domain labels. The random selection ratio of CSAL is set to 0.1, which is much lower than the setting of 0.3. Therefore, we provide here the experimental results of the on the T2 datasets with the labeling budgets set to 0.1.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | | |
|  | **FP** | **FN** | **delay** |
| T2S | 0 | 0 | 514 |
| T2G | 0.4 | 0.1 | 1304 |
| T2I | 0 | - | - |

The results in the table show that the delay and FP of the improve significantly after the label rate decreases. It indicates that the error rate-based detection method tends to be correlated with the active learning labeling budgets. Another reason is that the prediction confidence-based drift detection strategy of PMH-DDM is insensitive to the decision boundary drift on a fixed data distribution.

In summary, PMH-DDM can effectively detect concept drift on multi-source data streams. It performs well for the data distribution drift and has an advantage in incremental concept drift detection. The experimental accuracy results on both synthetic and real datasets shows that PMH-DDM has dedication to the learning performance of CSAL.

Due to the length of the manuscript, the analysis of PMH-DDM algorithm based on ablation study is given in Section 5.5. Since the concept drift positions are known for the synthetic datasets, the average drift detection results of PMH-DDM on T1G and T3S were also added to the Figure 3 of the synthetic datasets.

[1] Gama, J., Medas, P., Castillo, G., & Rodrigues, P. (2004). Learning with drift detection. In *Advances in Artificial Intelligence–SBIA 2004: 17th Brazilian Symposium on Artificial Intelligence, Sao Luis, Maranhao, Brazil, September 29-Ocotber 1, 2004. Proceedings 17* (pp. 286-295). Springer Berlin Heidelberg.

[2] Frias-Blanco, I., del Campo-Ávila, J., Ramos-Jimenez, G., Morales-Bueno, R., Ortiz-Diaz, A., & Caballero-Mota, Y. (2014). Online and non-parametric drift detection methods based on Hoeffding’s bounds. *IEEE Transactions on Knowledge and Data Engineering*, *27*(3), 810-823.

[3] Bifet, A., & Gavalda, R. (2007, April). Learning from time-changing data with adaptive windowing. In *Proceedings of the 2007 SIAM international conference on data mining* (pp. 443-448). Society for Industrial and Applied Mathematics.

[4] Yu, H., Liu, W., Lu, J., Wen, Y., Luo, X., & Zhang, G. (2023). Detecting group concept drift from multiple data streams. Pattern Recognition, 134, 109113.

[5] Gemaque, R. N., Costa, A. F. J., Giusti, R., & Dos Santos, E. M. (2020). An overview of unsupervised drift detection methods. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(6), e1381.