

Compsci 361 Assignment 6

Hasnain Cheena

190411106

In this assignment a dataset containing information about New South Wales Electricity has been run through two algorithms; the Hoeffding Tree and Adaptive Hoeffding Tree. The tool used to run the algorithms on the data was MOA.

The evaluation technique used on both algorithms is the prequential method. This method consists of using each instance in the stream to test model performance before using the instance to train the model with emphasis on more recent instances. The key parameters examined for discussion are leaf prediction, grace period and split confidence. The parameters are defined as:

1. Leaf prediction defines the mechanism used for prediction of the target label at the leaves. Three options are available; the majority class classifier, the Naïve Bayes classifier and the Adaptive Naïve Bayes classifier.
2. Grace period is the number of instances a leaf observes between splits.
3. Split Confidence is defined as the allowable error in a split.

To aid reproducibility, parameter settings were kept default unless the parameter was under consideration. Furthermore, runtime and accuracy were attributes used to monitor performance. Runtime was calculated by running the algorithm with the specified parameters ten times and then averaging to get a reliable runtime.

Question 1 – Hoeffding Tree

Initially, the Hoeffding Tree was run using all default MOA settings. The graph of this run is shown below in Figure 1. It can clearly be seen that the model performance is degrading over time while memory usage continues to soar. The model degradation indicates concept drift has occurred.

From the runs it can be observed that the mean accuracy and run time are sensitive to the parameters listed above. With the increase of split confidence, the mean accuracy and runtime of the Hoeffding Tree increases as shown in Table 1. Furthermore, as grace period increases the mean accuracy and runtime of the Hoeffding Tree decreases as shown in Table 2. Finally, looking at Table 3 when a majority class classifier is used as the leaf prediction mechanism on a Hoeffding Tree it produces the lowest mean accuracy. In comparison, Naïve Bayes and Adaptive Naïve Bayes produce better results (with Adaptive Naïve Bayes having a higher mean accuracy than Naïve Bayes). All three leaf predictors have similar runtimes.

Question 2 – Hoeffding Adaptive Tree

Initially, the Adaptive Hoeffding Tree was run using all default MOA settings. The graph of this run is shown below in Figure 1. Once more, concept drift can be seen occurring. However, unlike the Hoeffding Tree, the Hoeffding Adaptive Tree was able to account for the concept drift by switching to a new tree that is more accurate than the current one. This is observable in the accuracy plot in Figure 1 at places where the two algorithms diverge significantly. Overall, the Hoeffding Adaptive Tree performs better (has a higher mean accuracy) on this dataset but has a slightly longer runtime than the normal Hoeffding Tree.

From the runs it can be observed that with an increase in split confidence, the mean accuracy and runtime of the Hoeffding Adaptive Tree increases as shown in Table 1. Furthermore, as grace period increases the mean accuracy and runtime of the Hoeffding Adaptive Tree decreases as shown in Table 2. Finally, looking at Table 3 when a majority class classifier is used as the leaf prediction mechanism on a Hoeffding Adaptive Tree it has the quickest runtime but produces the lowest mean accuracy. In comparison, Naïve Bayes and Adaptive Naïve Bayes produce better results (with Adaptive Naïve Bayes having a higher mean accuracy than Naïve Bayes) with a slightly longer runtime.

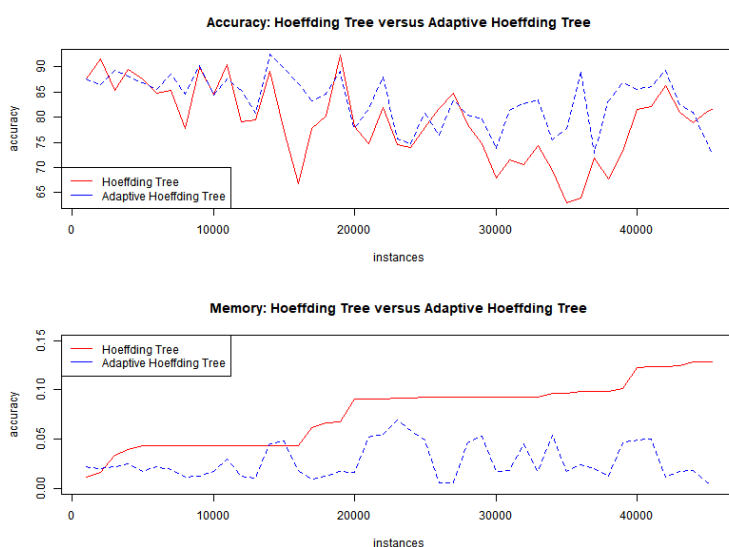


Figure 1: Plots of Accuracy and Memory

Table 1: Split Confidence

Split Confidence	Mean accuracy (%) Hoeffding Tree	Run time (s) Hoeffding Tree	Mean accuracy (%) Hoeffding Adaptive Tree	Run time (s) Hoeffding Adaptive Tree
0	79.21	0.83	82.23	0.95
0.05	82.54	1.33	85.78	1.36
0.1	82.90	1.64	86.28	1.34
0.5	82.78	2.14	86.42	1.52
0.8	82.78	2.16	86.42	1.55

Table 2: Grace Period

Grace period	Mean accuracy (%) Hoeffding Tree	Run time (s) Hoeffding Tree	Mean accuracy (%) Hoeffding Adaptive Tree	Run time (s) Hoeffding Adaptive Tree
50	79.52	0.97	84.59	1.23
200	79.21	0.77	83.23	1.08
500	79.82	0.64	83.50	1.03
1000	78.91	0.62	82.31	0.88
10000	75.83	0.41	76.64	0.55

Table 3: Leaf prediction

Leaf Predictor	Mean accuracy (%) Hoeffding Tree	Run time (s) Hoeffding Tree	Mean accuracy (%) Hoeffding Adaptive Tree	Run time (s) Hoeffding Adaptive Tree
Majority Class	75.86	0.62	75.62	0.91
Naïve Bayes	77.35	0.61	82.53	1.09
Adaptive Naïve Bayes	79.21	0.64	83.23	1.09