MOA Session Lab Report

M2 D&K Xiaoxuan HEI

1. Classifiers

- NaiveBayes
- SAMkNN
- AdaptiveRandomForest
- HoeffdingTree
- DriftDetectionMethodClassifier

2. Results

Among these 5 classifiers, SAMkNN and AdaptiveRandomForest has the highest accuracy. But their execution time is much longer than the other three. Especially the SAMkNN, its execution time is up to 16 min. The accuracy of NaiveBayes is the lowest.
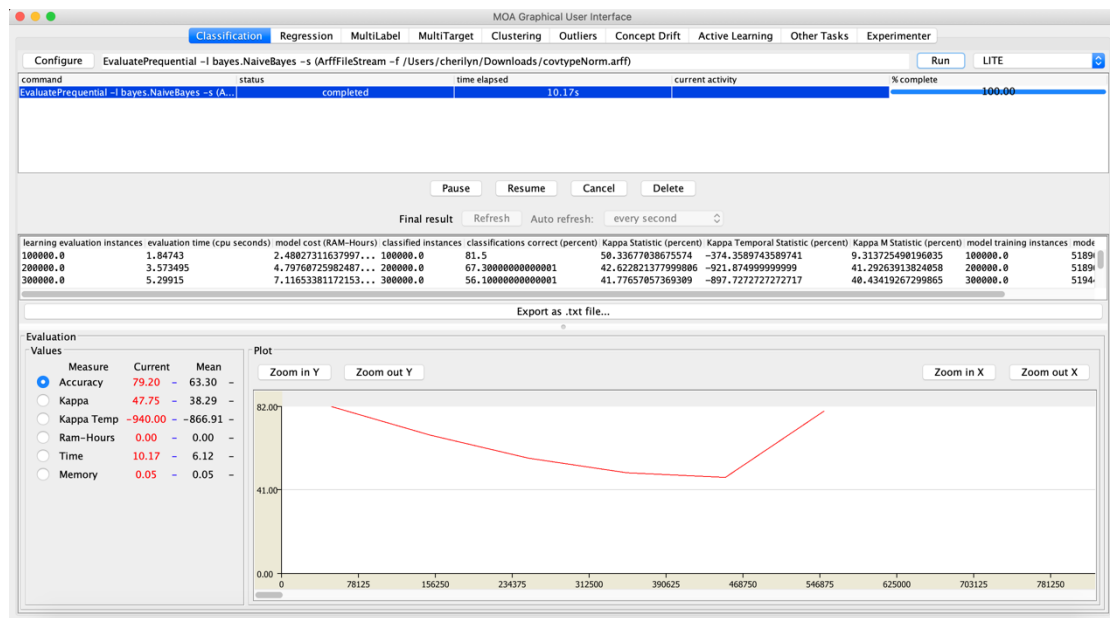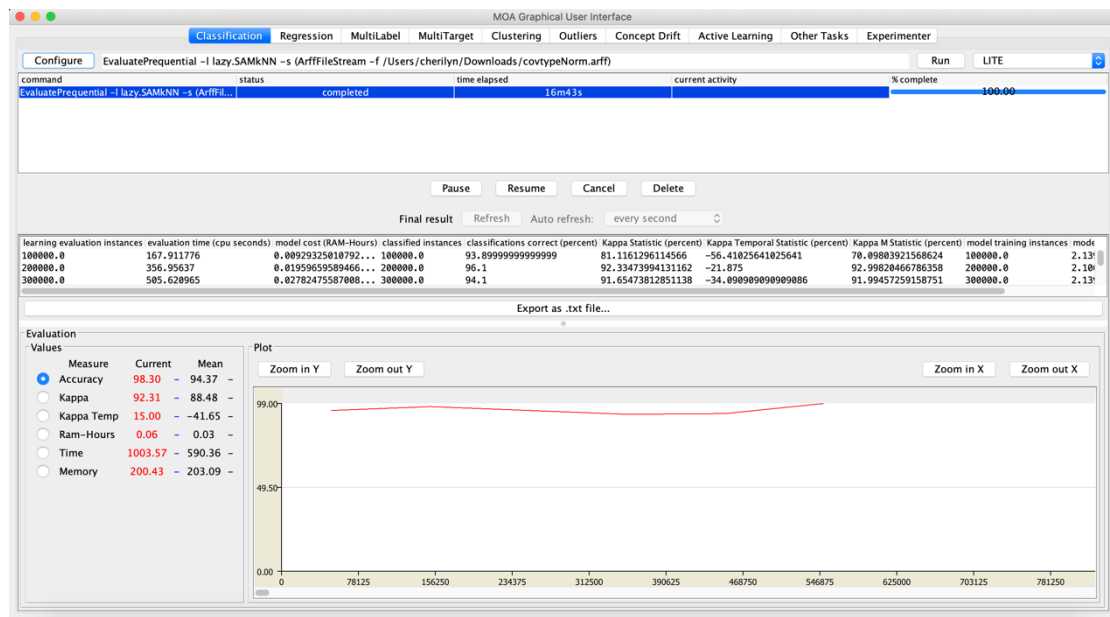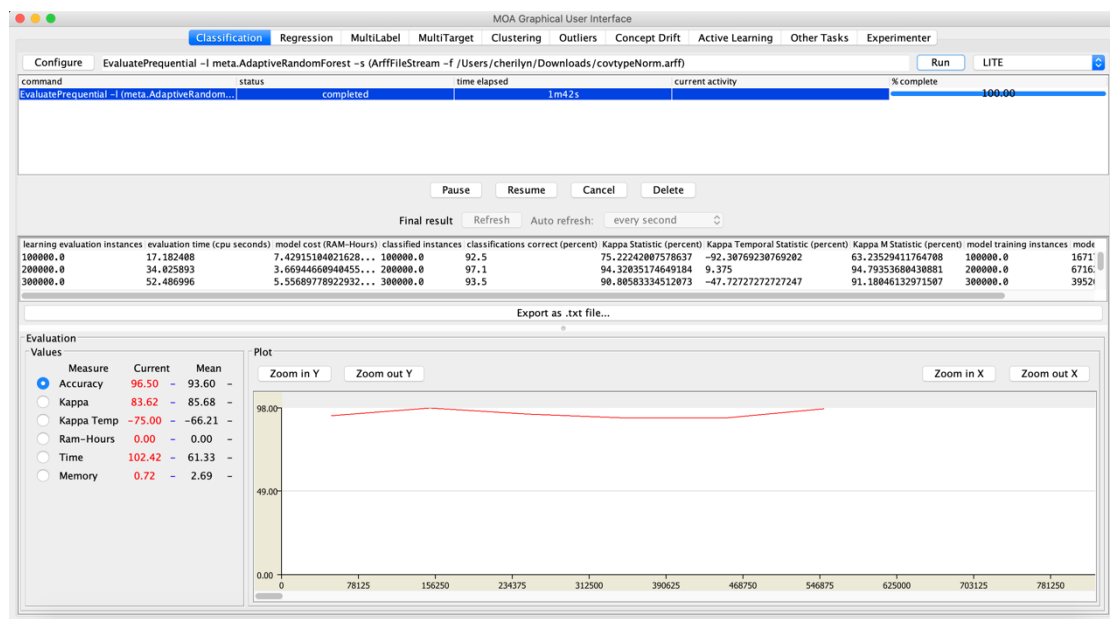

Figure 1 - NaiveBayes

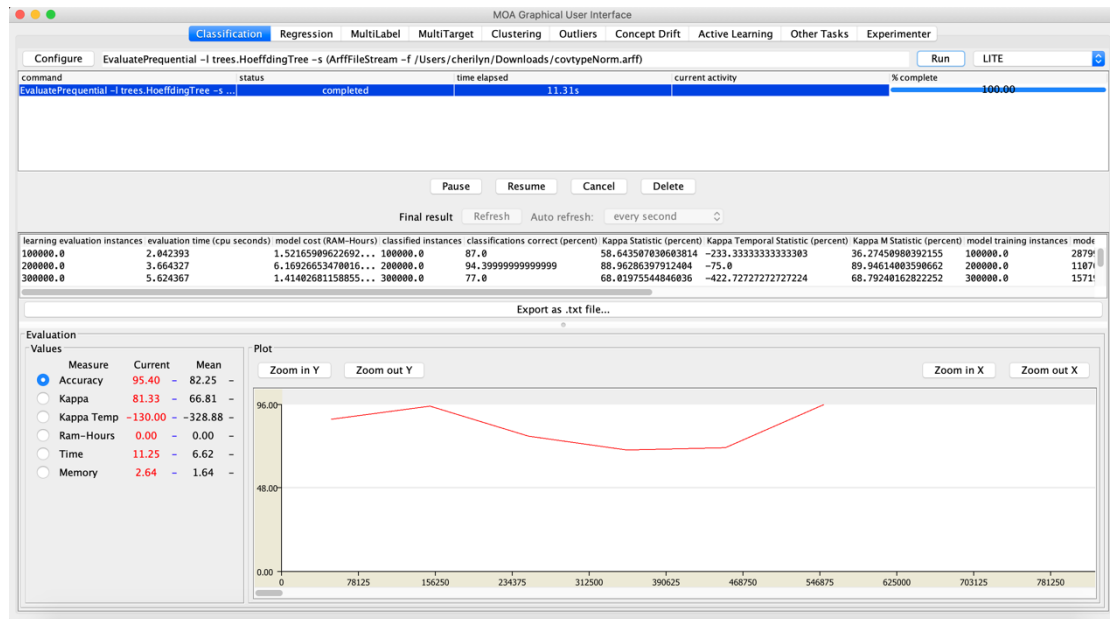*Figure 2 - SAMkNN*

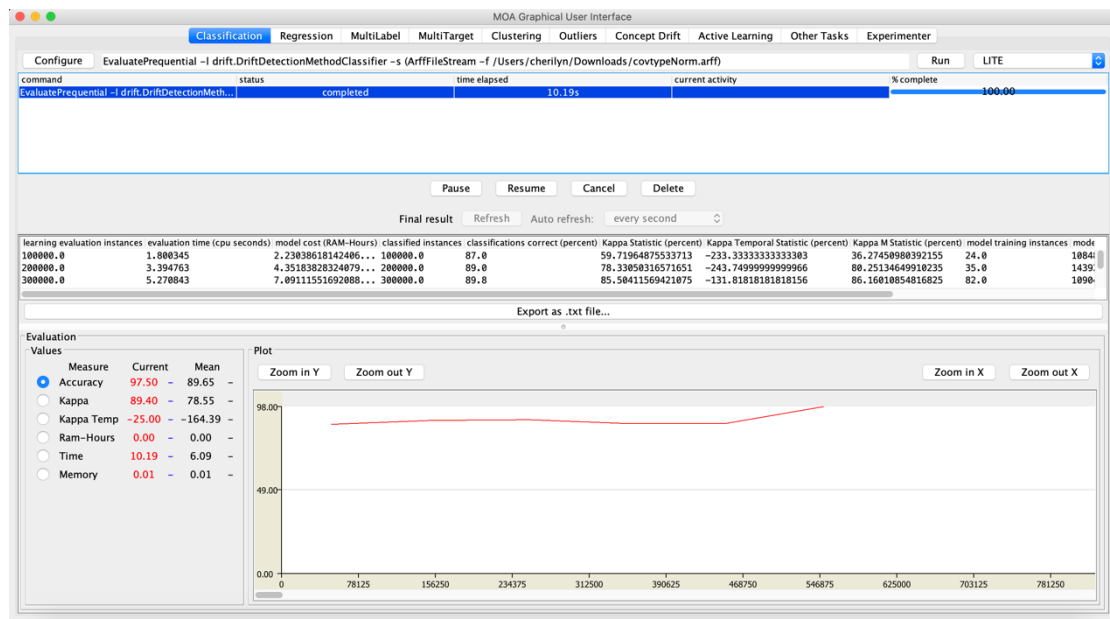*Figure 3 - AdaptiveRandomForest*

Figure 4 - HoeffdingTree

Figure 5 - DriftDetectionMethodClassifier

3. Discussion about the results

- NaiveBayes

In the forest coverage type dataset, many columns have dependencies. The Naive Bayes classifier is based on the independence between attributes, so this classifier performed not so well.

From Figure 1, we can learn that the average accuracy of the Naive Bayes classifier is only 63.30%.

- Although SAMkNN has the highest accuracy, its execution time and memory are unacceptable.
- As for Hoeffding Tree, there is nothing special. Its performance on accuracy, execution time and memory are not so good, and its curve is not very smooth.
- AdaptiveRandomForest vs DriftDetectionMethodClassifier

The red line represents Adaptive Random Forest and the blue line represents Drift Detection Method Classifier. We can find that the accuracy of ARF is slightly higher than DDM, while the curve of DDM is smoother. And the execution time and memory usage of DDM are much less than that of ARF. But the execution time and memory usage of ARF are still acceptable, so if you have high requirements on accuracy, you can choose ARF
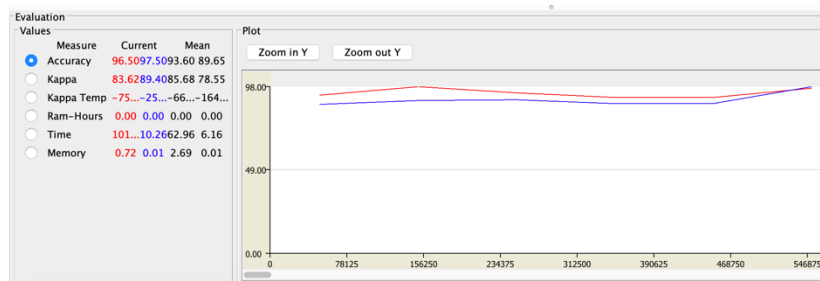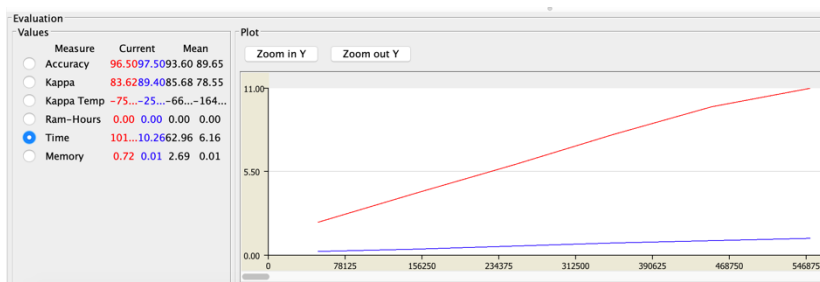

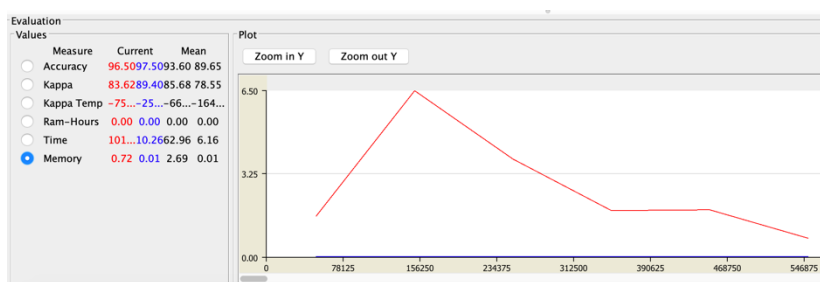*Figure 6 - Comparision of accuracy*

Figure 7 - Comparision of time

Figure 8 - Comparision of memory

4. Recommendation

According to the performance and the discussion above, considering accuracy, time and memory, I recommend DriftDetectionMethodClassifier.