Learning Decision Trees with Reinforcement Learning: Supplementary Material

A. Complete Experimental Results

Figure 1 demonstrates that the AUC score of RLBDT on the validation set improves over time and exceeds the AUC score of CART on all the datasets.

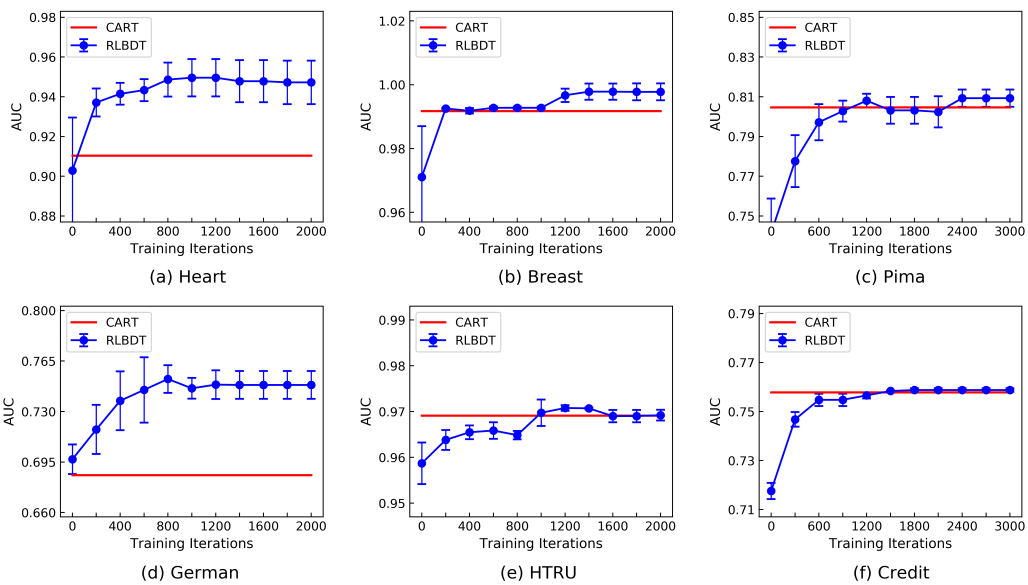

Figure 1: The AUC score of RLBDT and CART on the validation set.

Figure 2 shows the AUC score of RLBDT and CART on the test set. The results imply that our method outperforms the CART baseline on all the datasets.

B. Design of the Reward Signal

In the experiment, we find that if we directly use the performance score on the validation set as the reward signal, the controller may find decision trees which perform extremely well on the validation set but poorly on the test set. This problem is more significant on datasets with relatively small sample number than on datasets with large sample number. A possible explanation to this phenomenon is that as the sample number decreases, the variance of the performance score on the validation set will increase, which makes the reward signal more noisy and leads the controller to overfit the validation set. To solve this problem, we design two methods to improve the quality of the reward signal.

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

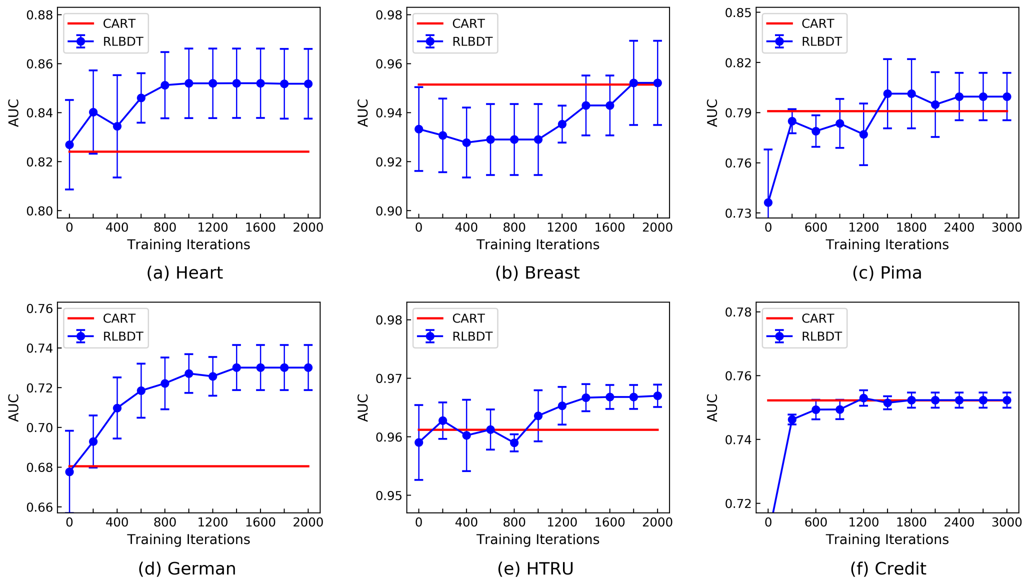

Figure 2: The AUC score of RLBDT and CART on the test set.

Firstly, we randomly choose 80% of the samples in the validation set and calculate the performance score on this subset. We repeat this process for 5 times and use their mean value as the final score, which can help reduce the variance of the performance score on the validation set.

Secondly, as a model will be more likely to perform well on unseen test data if it has achieved good performance on both the training and validation set, we design the reward signal R as the F1 score of the performance scores on both sets:

$$R = \frac{2 \times PV \times PT}{PV + PT},$$

where PT is the performance score on the training set, and PV is the performance score on the validation set calculated in the way presented in the last paragraph.