Learning Decision Trees with Reinforcement Learning: Supplementary Material

Due to the page limit of workshop paper, only a part of the experimental results are illustrated in the main body. This supplementary material demonstrates the complete experimental results.

A. Results of Experiment 1

Experiment 1 has been introduced in the paper. All the datasets are split into training, validation and test set in proportion of $50\%, 25\%, 25\%$. Table 1 shows the final AUC score of RDT and CART on the validation and test set.

Table 1: Final AUC score of RDT and CART.

Datasets	Validation Set			Test Set			Search	Iters
	CART	RDT Avg	RDT Top	CART	RDT Avg	RDT Top	Space	
Pima	80.47	$\mathbf{80.93 \pm 0.44}$	81.54	79.09	$\mathbf{79.96 \pm 1.70}$	78.90	10^{13}	3000
Heart	91.04	$\mathbf{94.73 \pm 1.09}$	96.98	82.41	$\mathbf{85.18 \pm 1.71}$	87.19	10^{19}	2000
Breast	99.18	$\mathbf{99.78 \pm 0.27}$	99.99	95.15	$\mathbf{95.22 \pm 2.06}$	95.93	10^{45}	2000
German	68.59	$\mathbf{74.84 \pm 0.96}$	76.19	68.05	$\mathbf{73.01 \pm 1.36}$	74.43	10^{42}	2000
HTRU	96.91	$\mathbf{96.92 \pm 0.17}$	97.09	96.12	$\mathbf{96.70 \pm 0.23}$	97.03	10^{13}	2000
Credit	75.79	$\mathbf{75.88 \pm 0.07}$	75.98	75.22	$\mathbf{75.23 \pm 0.29}$	75.44	10^{42}	3000

Figure 1 shows that RDT's ROC score on the validation set improves over time and exceeds CART baseline on all the datasets.

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

B. Results of Experiment 2

In experiment 2, the samples in each dataset are shuffled randomly and split into training, validation and test set in proportion of $60\%, 20\%, 20\%$. This experiment is designed to test the reproducibility of the results in experiment 1. The final AUC score of RDT and CART on the validation and test set is shown in Table 2.

Table 2: Final AUC score of RDT and CART.

Datasets	Validation Set			Test Set			Search	Iters
	CART	RDT Avg	RDT Top	CART	RDT Avg	RDT Top	Space	
Pima	84.46	$\mathbf{87.14 \pm 1.16}$	89.10	73.61	$\mathbf{76.22 \pm 1.96}$	77.84	10^{13}	3000
Heart	87.29	$\mathbf{93.46 \pm 0.66}$	93.89	85.62	$\mathbf{87.35 \pm 2.04}$	89.34	10^{19}	2000
Breast	96.72	$\mathbf{98.91 \pm 0.58}$	99.61	94.51	$\mathbf{95.61 \pm 1.52}$	97.52	10^{45}	2000
German	73.08	$\mathbf{77.97 \pm 0.95}$	79.31	69.18	$\mathbf{71.90 \pm 1.22}$	73.60	10^{42}	2000
HTRU	97.55	$\mathbf{97.56 \pm 0.04}$	97.61	**97.36**	97.34 ± 0.00	97.34	10^{13}	2000
Credit	75.89	$\mathbf{75.93 \pm 0.14}$	76.14	**75.10**	74.74 ± 0.05	74.68	10^{42}	3000

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

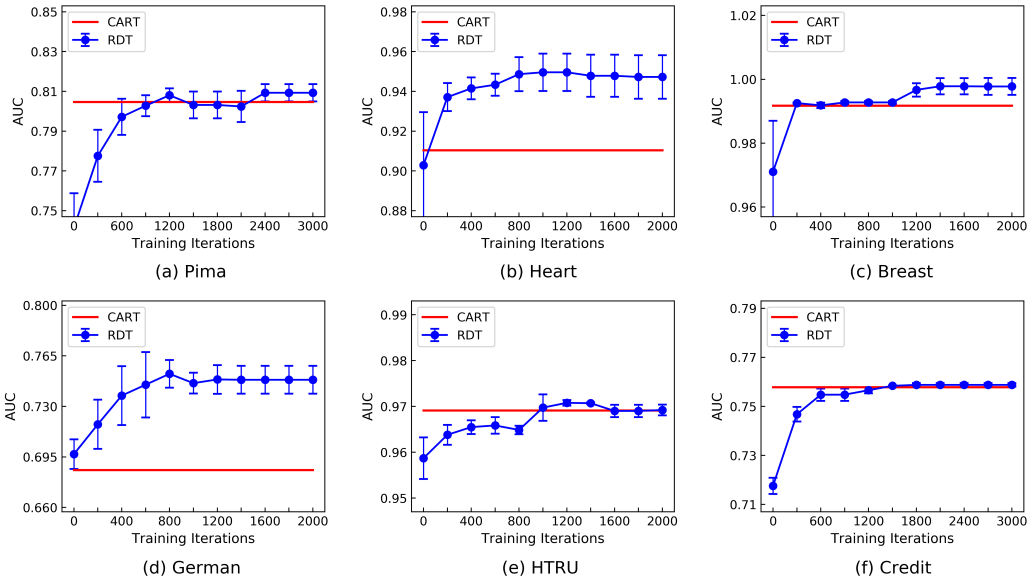

Figure 1: AUC score on the validation set of RDT and CART in experiment 1.

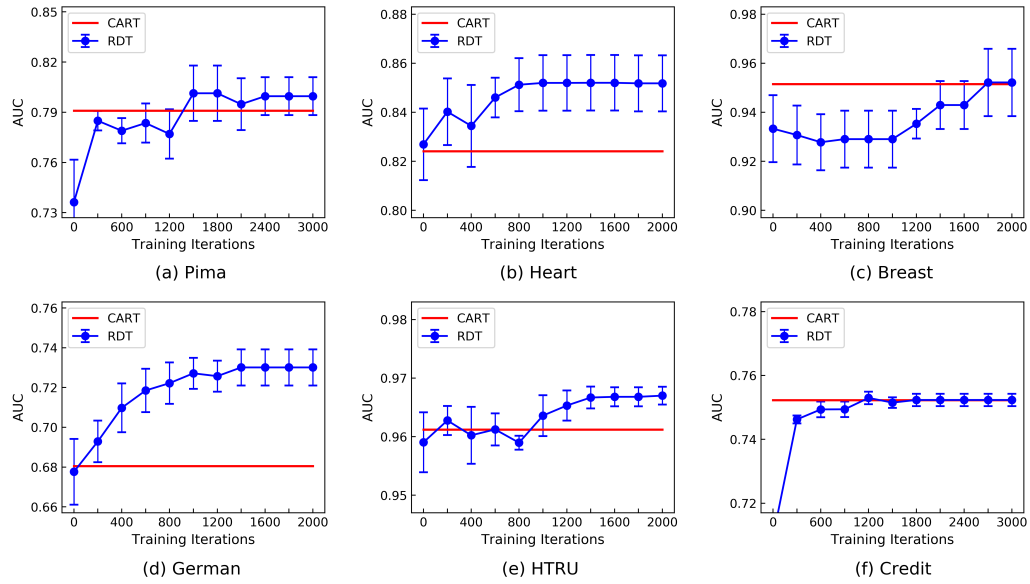

Figure 2: AUC score on the test set of RDT and CART in experiment 1.

Figure 3 and Figure 4 demonstrate AUC score of RDT and CART on the validation and test set respectively. The results demonstrate that RDT acquires competitive performance compared to CART, which is consistent with the conclusion in experiment 1 and validates the effectiveness of our method.

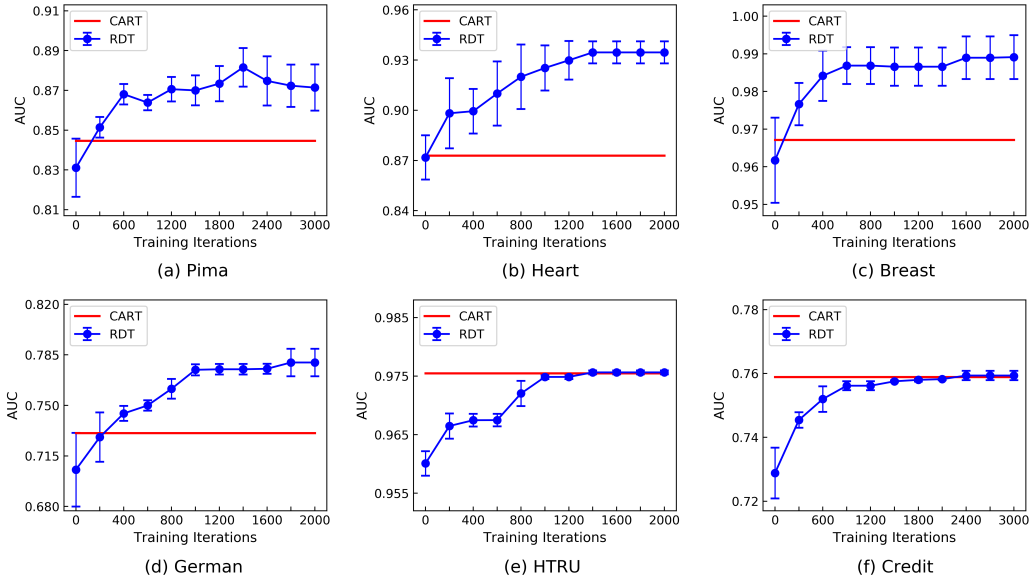

Figure 3: AUC score on the validation set of RDT and CART in experiment 2.

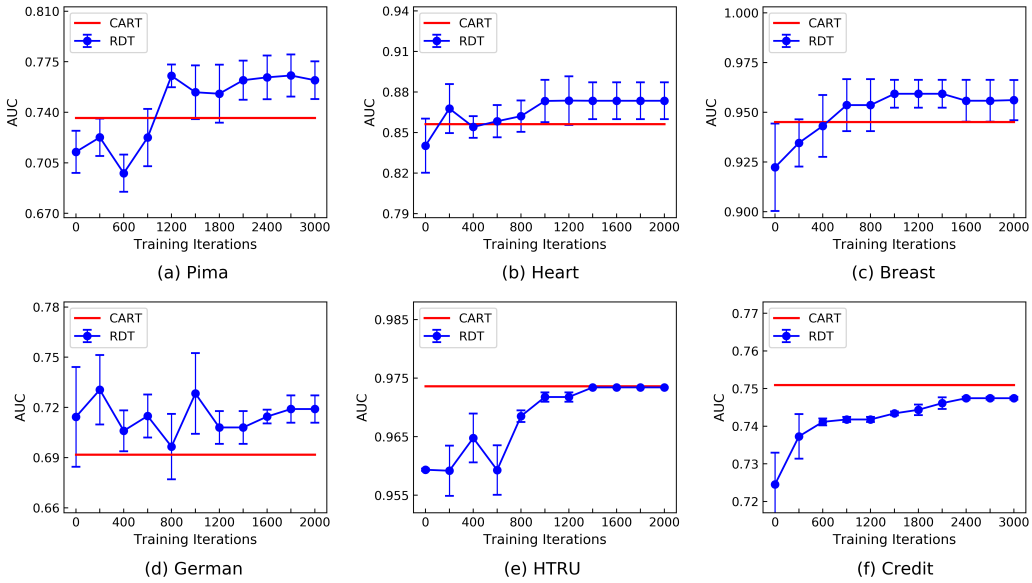

Figure 4: AUC score on the test set of RDT and CART in experiment 2.

C. Design of Reward Signal

In the experiment, we find that if we directly use the performance score on the validation set as 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 large datasets. 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 on the validation set. To solve this problem, we design two methods to improve the quality of the reward signal.

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

$$R = \frac{2 \times PV \times PT}{PV + PT} \tag{1}$$

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.