Citation: Feng-fei Zhao, Zheng Qin, Zhuo Shao, *et al*., 2014. Greedy feature replacement for online value function approximation. *Journal of Zhejiang University-Science C (Computers & Electronics)*, 15(3):223-231. [doi:10.1631/jzus.C1300246]

Greedy feature replacement for online value function approximation

用于在线值函数近似的贪婪特征替换方法

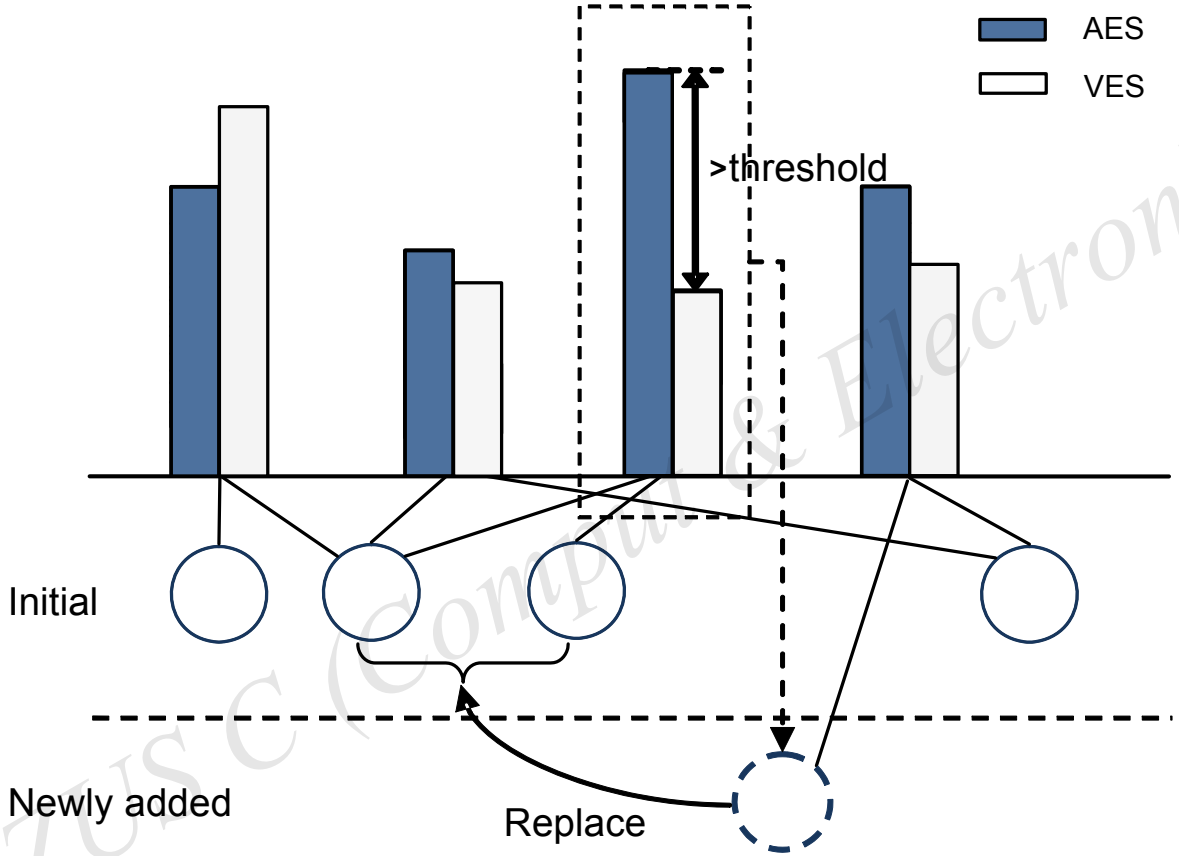
Key words: Reinforcement learning, Function approximation, Feature dependency, Online expansion, Feature replacement

关键词：强化学习；值函数近似；特征依赖；在线扩展；特征替换

1. Reinforcement learning (RL) in real-world problems requires function approximations that depend on selecting the appropriate feature representations. Representational expansion techniques can make linear approximators represent value functions more effectively; however, most of these techniques function well only for low dimensional problems.

2. In this paper, we present the greedy feature replacement (GFR), a novel online expansion technique, for value-based RL algorithms that use binary features. Given a simple initial representation, the feature representation is expanded incrementally. New feature dependencies are added automatically to the current representation and conjunctive features are used to replace current features greedily. The virtual temporal difference (TD) error is recorded for each conjunctive feature to judge whether the replacement can improve the approximation. Correctness guarantees and computational complexity analysis are provided for GFR.

The principle of greedy feature replacement (GFR)



Initial features are represented by solid circles and newly added features by dotted circles. The bar chart tracks the accumulation of approximation errors for each simultaneously activated feature pair. If the AES (approximation error sum) is higher than the VES (virtual error sum) and the difference reaches the user-defined threshold, feature replacement is performed

Experimental results of greedy feature replacement (GFR)

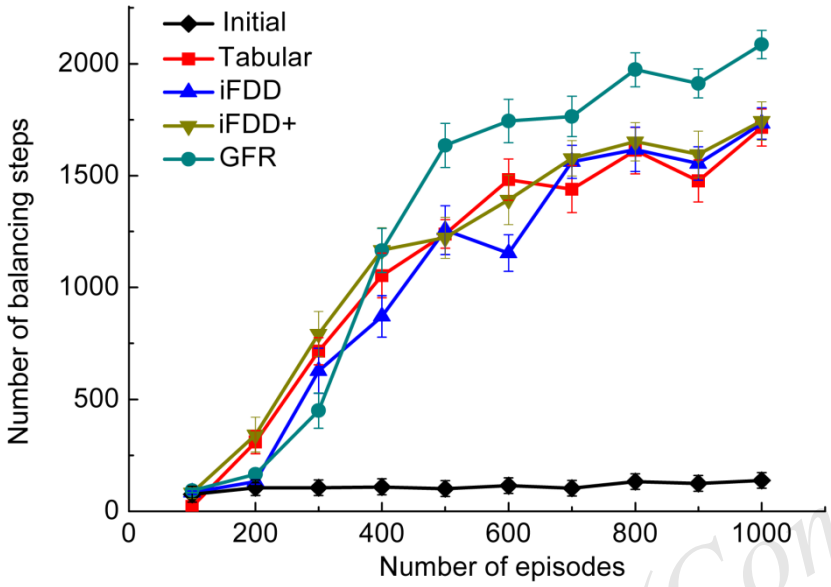


Fig. 3 The experimental results on the inverted pendulum domain, averaged over 30 independent trials

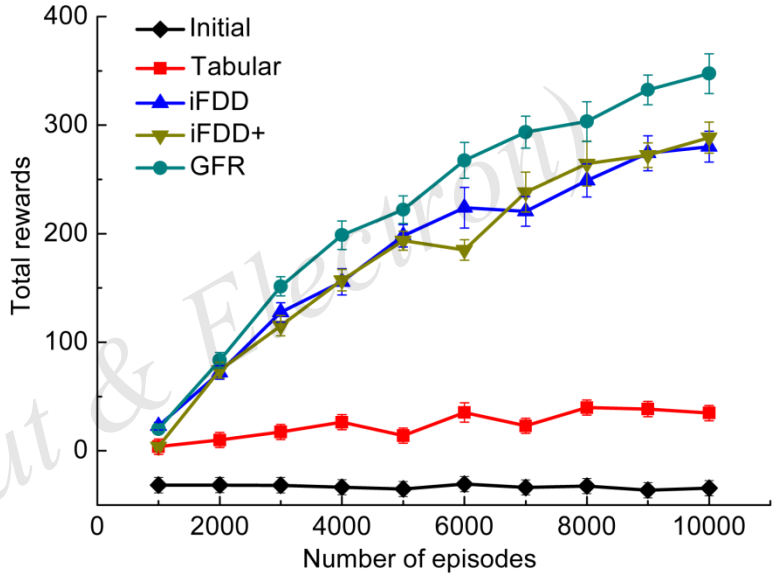


Fig. 5 The experimental results on the persistent surveillance domain, averaged over 30 independent trials

The proposed method learns much faster and has advantages in solving large-scale problems.