International Journal of Innovative Computing, Information and Control
Volume **5**, Number **4**, April **2009**

 ISSN 1349-4198
pp. 913–922

CONVERGENCE ANALYSIS ON TEMPORAL DIFFERENCE LEARNING

Jinsong Leng[1], Lakhmi Jain[1] and Colin Fyfe[2]

[1]School of Electrical and Information Engineering
University of South Australia
Mawson Lakes SA 5095, Australia
{ Jinsong.Leng; Lakhmi.Jain }@unisa.edu.au

[2]Applied Computational Intelligence Research Unit
University of the West of Scotland
colin.fyfe@paisley.ac.uk

Received September 2007; revised February 2008

Abstract. *Learning to act in an uncertain environment without external instruction is considered as one of the fundamental features of intelligence. Temporal difference (TD) learning is an incremental learning approach and has been widely used in various application domains. Utilising eligibility traces is an important mechanism in enhancing learning ability. For large, stochastic and dynamic systems, however, the TD method suffers from two problems: the state space grows exponentially with the curse of dimensionality and there is a lack of methodology to analyse the convergence and sensitivity of TD algorithms. Measuring learning performance and analysing sensitivity of parameters are very difficult and expensive, and such performance metrics are obtained only by running an extensive set of experiments with different parameter values. In this paper, convergence is investigated by performance metrics, which is obtained through simulating a game of soccer. Sarsa(λ) learning control algorithm, in conjunction with a linear function approximation technique known as tile coding, is used to help soccer agents learn the optimal control processes. This paper proposes a methodology for finding the optimal parameter values to improve the quality of convergence.*
Keywords: Temporal difference learning, Agent, Convergence analysis

1. **Introduction.** Most agent-based systems are considered as real-time, stochastic and dynamic environments, and can be modeled as Markov Decision Processes (MDPs) [1]. An agent can be defined as a hardware and/or software-based computer system displaying the properties of autonomy, social adaptness, reactivity, and proactivity [2]. Informally, learning is considered as the process of acquiring new knowledge and cognitive skills by incorporating the prior knowledge and skills that lead to some improvements in performance. Dynamic programming (DP) [3] and heuristic search [4] are full backups techniques that work well in small, stochastic and deterministic systems.

For large and non-deterministic environments, it is often extremely complex and difficult to formally verify their properties a *priori*. An agent has to learn what action to take by interacting with the environment without external instruction. Reinforcement learning (RL) is the learning for helping agents pursue the goals and find the optimal action at a given state, so as to maximise numerical reward from and while interacting with the environment. The TD method [5] has become the most popular RL technique.

The performance of RL algorithms is sensitive not only to the generation of the approximation function, but also to the parameter values. Consequently, the performance of TD algorithms depends on the way the state space and the values of the parameters are