# Reinformcement Learning Theory Learning

Yue Wang

*MSRA*

---

**Abstract**

a simple note for RL theroy

*Keywords:*

---

## 1. Introduction

RL theory is very long history and these days

- Bullet point one
- Bullet point two

1. Numbered list item one
2. Numbered list item two

## 2. Notation and Formatting

some notations:

| | |
|---|---|
| $\mathcal{S}$ | the state space |
| $S_t$ | the state at time t, stochastic |
| $s_t$ | the state at time t, actual |
| $s$ | the state, actual |
| $R_t$ | the reward at time t, stochastic |
| $r_t$ | the reward at time t, actual |
| $r$ | the reward, actual |

## 3. RL algorithms

## 4. RL convergence theory

### *4.1. counterexample*

In Sutton and Barto [1, chap 11.3] the authors give an intuitive conclusion about when these algorithms will divergence :

> the danger of instability and divergence arises whenever we combine three things:
>
> 1. training on a distribution of trainsition other than that naturally generated by the process whose expectation is being estimated(e.g. off-policy learning)
> 2. scalable function approximation (e.g. li)

#### *4.1.1. counterexample1*

In Tsitsiklis and Roy [2] Dimitri [3]

$E[\theta_(t+1)|\theta_t]$

## References

[1] R. S. Sutton, A. G. Barto, Reinforcement learning: An introduction, 2011.

[2] J. Tsitsiklis, B. V. Roy, Feature-based methods for large scale dynamic programming, Machine Learning (1996).

[3] P. Dimitri, A Counterexample to Temporal Differences Learning, pdfs.semanticscholar.org (????).