

TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI  
VIỆN TOÁN ỨNG DỤNG VÀ TIN HỌC

---

MÔ HÌNH HÓA DỰA TRÊN CÁ THỂ  
MÔ HÌNH RẦY NÂU HẠI LÚA

Nghiên cứu ảnh hưởng phân bố không gian của hoa  
thu hút thiên địch rầy lên sự phát triển của rầy nâu

ĐỒ ÁN TỐT NGHIỆP

Chuyên ngành: TOÁN TIN

Giáo viên hướng dẫn: TS. NGUYỄN THỊ NGỌC ANH

Sinh viên thực hiện: VŨ THU THẢO

Lớp: Toán Tin 2 - K55

HÀ NỘI - 2015

# Mục lục

<b>Mở đầu</b>	<b>3</b>
0.1 Giới thiệu . . . . .	4
0.2 Qua trình quyết định Markov . . . . .	4
<b>Danh mục công trình công bố của tác giả</b>	<b>8</b>

# Bài toán

# Phương pháp

## 0.1 Giới thiệu

Trong ngành khoa học máy tính, học tăng cường (Reinforcement Learning) là một lĩnh vực con của học máy (Machine Learning), nghiên cứu cách thức một tác tử (agent) trong một môi trường nên chọn thực hiện các hành động nào để có được phần thưởng có giá trị lớn nhất về lâu về dài. Các thuật toán học tăng cường cố gắng tìm một chiến lược ánh xạ từ không gian trạng thái tới không gian hành động mà agent nên chọn trong các trạng thái đó.

## 0.2 Qua trình quyết định Markov

Quá trình quyết định Markov (Markov Decision Processes, ký hiệu là MDP) cung cấp một nền tảng toán học cho việc mô hình hóa việc ra quyết định trong các tình huống mà kết quả là một phần ngẫu nhiên, một phần dưới sự điều khiển của một người ra quyết định. MDP rất hữu dụng trong việc học một loạt các bài toán tối ưu hóa được giải quyết thông qua quy hoạch động và học tăng cường. MDP được biết đến sớm nhất vào những năm 1950.

MDP là một quá trình điều khiển ngẫu nhiên thời gian rời rạc. MDP là một tập 5 dữ liệu  $\langle \mathbb{S}, \mathbb{A}, P, R, \gamma \rangle$ . Trong đó:

$\mathbb{S}$ : không gian các trạng thái

$$\mathbb{P} : \mathbb{S} \times \mathbb{A} \times \mathbb{S} \rightarrow$$

$$R : \mathbb{S} \times \mathbb{A} \rightarrow$$

giả sử rằng, tại thời điểm  $t$ , trạng thái  $S_t = s$  và agent có hành động  $A_t = a$ . khi đó, xác suất của trạng thái  $B \in \mathbb{S}$  tại thời điểm  $t + 1$  được cho bởi công thức:

$$P(s, a, B) = \mathbb{P}(S_{t+1} | S_t = s, A_t = a) \quad (1)$$

sau quá trình này, agent nhận được một phần thưởng ngẫu nhiên là  $R(t+1)$ . Hàm thưởng  $R(s, a)$  là phần thưởng thu được khi thực hiện hành động  $a$  ở trạng thái  $s$

$$R(s, a) = \mathbb{E}[R(t+1)|S_t = s, A_t = a] \quad (2)$$

Tại bất kì bước thời gian nào, agent chọn hành động của nó theo một chính sách  $\pi : \mathbb{S} \times A \rightarrow \mathbb{R}$  sao cho với mỗi  $s \in \mathbb{S}, C \rightarrow \pi(s, C)$  là xác suất phân phối trên  $(A, A)$ . Do đó, chính sách  $\pi$  và trạng thái ban đầu  $s_0 \in \mathbb{S}$  xác định chuỗi trạng thái - hành động - phần thưởng ngẫu nhiên  $(S_t, A_t, R(t+1))_{t \geq 0}$  với giá trị trên  $\mathbb{S} \times A \times \mathbb{R}$ . Trong một không gian vô hạn, hiệu suất của agent thường được tính bằng tổng phần thưởng chiếu khấu thu được sau một chính sách là

$$G_t = \sum_{k=0}^{\infty} \gamma^k R(t+k+1) \quad (3)$$

Vì phần thưởng này là ngẫu nhiên, agent xem xét giá trị kì vọng của nó, thường được gọi là state - value function

$$V_\pi(s) = \mathbb{E}_\pi[G_t | S_t = s] \quad (4)$$

trong đó, chỉ số  $\mathbb{E}_\pi$  chỉ ra rằng xấp xỉ hành động được chọn theo chính sách  $\pi$ . State - value function được đo tốt khi agent ở trong một trạng thái nhất định và tuân theo một chính sách nhất định. tương tự, ta có Action - value function

$$Q_\pi(s, a) = \mathbb{E}_\pi[G_t | S_t = s, A_t = a] \quad (5)$$

ta có mối liên hệ giữa  $V_\pi$  và  $Q_\pi$

$$V_\pi(s) = \int_A \pi(s, a) Q_\pi(s, a) da \quad (6)$$

Hầu như tất cả các thuật toán học tăng cường được thiết kế để tính các hàm giá trị này dựa trên các phương trình Bellman

$$V_\pi(s) = R_\pi(s) + \gamma T_\pi V_\pi(s) \quad (7)$$

$$Q_\pi(s, a) = R(s, a) + \gamma T_a V_\pi(s) \quad (8)$$

khi chúng ta biểu diễn bởi  $T_a$  thì toán tử chuyển với hành động  $a$  (chính sách  $\pi$ )

$$T_s F(s) = \mathbb{E}[F(S(t+1)) | S_t = s, A_t = a] = \int_S P(s, a, s') F(s') ds' \quad (9)$$

$$T_\pi F(s) = \mathbb{E}_\pi[F(S(t+1)) | S_t = s] = \int_A \pi(s, a) \int_S P(s, a, a') F(s') F(s') ds' da \quad (10)$$

# Ứng dụng

## Kết luận

# Danh mục công trình công bố của tác giả

- [1] Vũ Thu Thảo, Nguyễn Ngọc Doanh, Nguyễn Nhị Gia Vinh, Nguyễn Thị Ngọc Anh, "*Mô hình hóa ảnh hưởng phân bố không gian của hoa diệt rầy tới sự phát triển của rầy nâu hại lúa*", Tạp chí Khoa học và Công nghệ - Đại học Thái Nguyên, 2015, pp.119-123.