AlphaOne Project

MaSSP 2018 – Computer Science

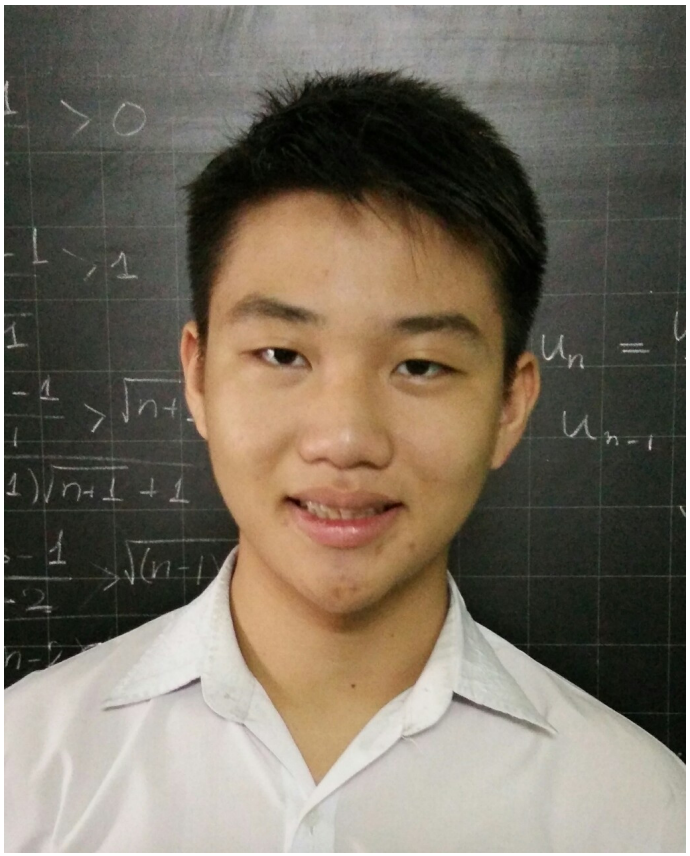
July 6, 2018

Team members

- Nguyễn Hoàng Hải - THPT Chuyên Nguyễn Bỉnh Khiêm, Quảng Nam.

Team members

- Nguyễn Quang Minh - Phổ thông Năng khiếu, ĐHQG TP.Hồ Chí Minh.



Team members

- Nguyễn Đức Thắng - THPT Chuyên Lê Quý Đôn, Vũng Tàu.

Supervisor

- Ngô Quốc Hưng - Universität Stuttgart (Germany)
  – Head mentor môn Tin MaSSP 2018

World Go Champion Defeated by AlphaGo

Motivation

1. Máy tính có thể vượt qua giới hạn con người!
2. Máy tính học để đưa ra quyết định
   $\longrightarrow$ Reinforcement Learning

Motivation

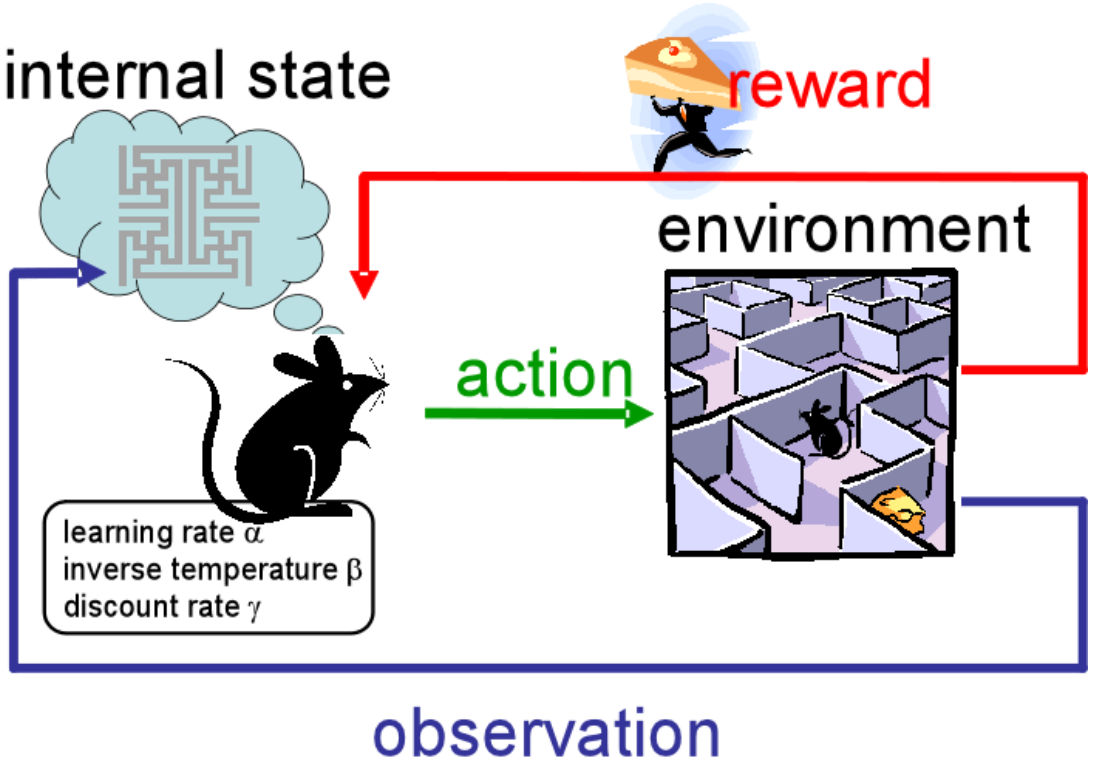
1. Máy tính có thể vượt qua giới hạn con người!
2. Máy tính học để đưa ra quyết định
   $\longrightarrow$ Reinforcement Learning

Những lĩnh vực cần xác định hành vi / ra quyết định:

1. Robotics
2. Self-driving cars
3. Games
4. Scheduling in airports, hospitals, etc.

What is Reinforcement Learning?

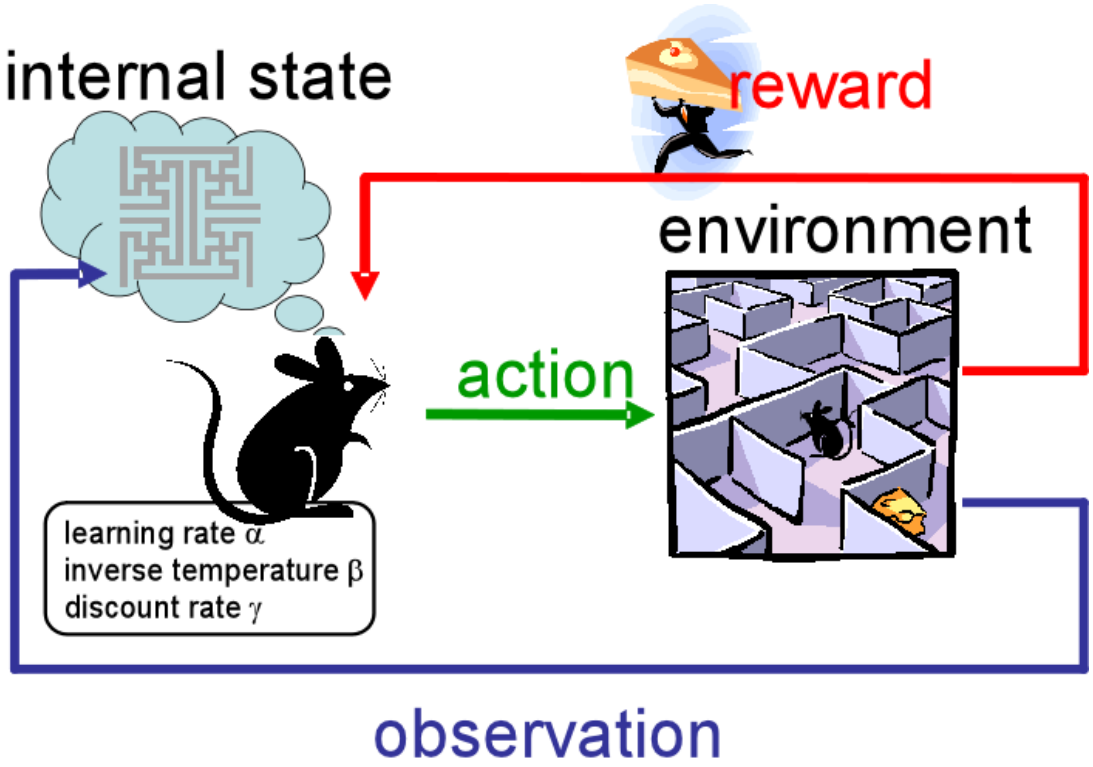
- Reinforcement Learning = Learning to Act (Học điều khiển)



(Source: The very basics of Reinforcement Learning)

Xác định hành vi dựa trên hoàn cảnh để đạt được lợi ích cao nhất.

Markov Decision Process



Markov Decision Process (MDP) = 5-tuple $\langle S, A, T, R, \gamma \rangle$

AlphaOne Project

Goal: dùng RL để huấn luyện cho máy tính chơi games (*Flappy Bird*, Gomoku cờ caro, Mario)

Planning đối với big model không hiệu quả $\rightarrow$ Learning

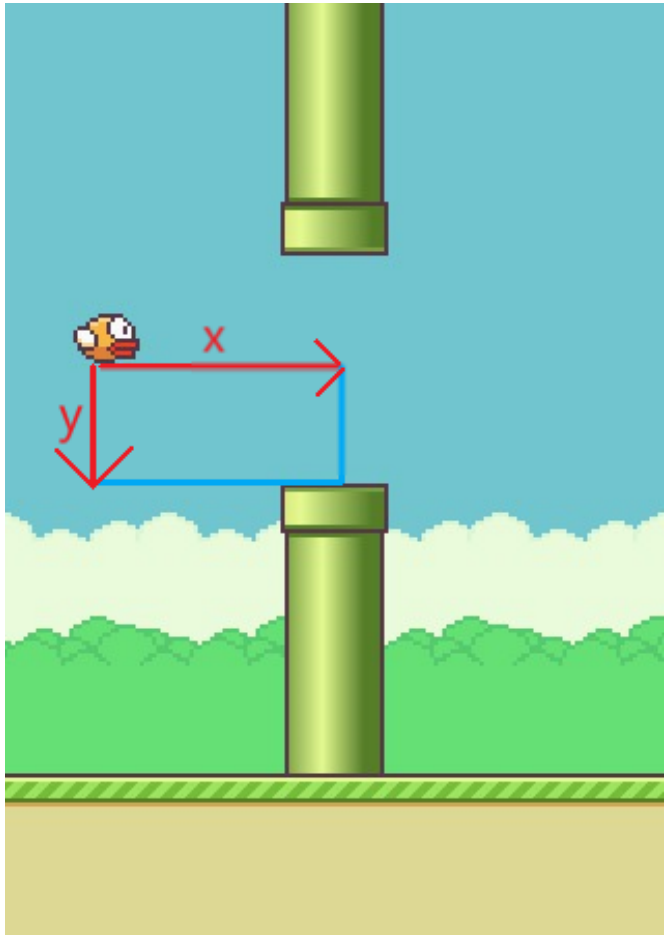
Model: Rewards và Transitions

- Hệ thống reward được sử dụng:
  - $r = 1$ nếu chim còn sống sau mỗi bước.
  - $r = -1000$ nếu chim chết.
- Transitions: Deterministic.
- Update equation for Q-learning:

$$Q_{i+1}(s, a) = Q_i(s, a) + \alpha(r + \gamma \max_{a'} Q_i(s', a') - Q_i(s, a))$$

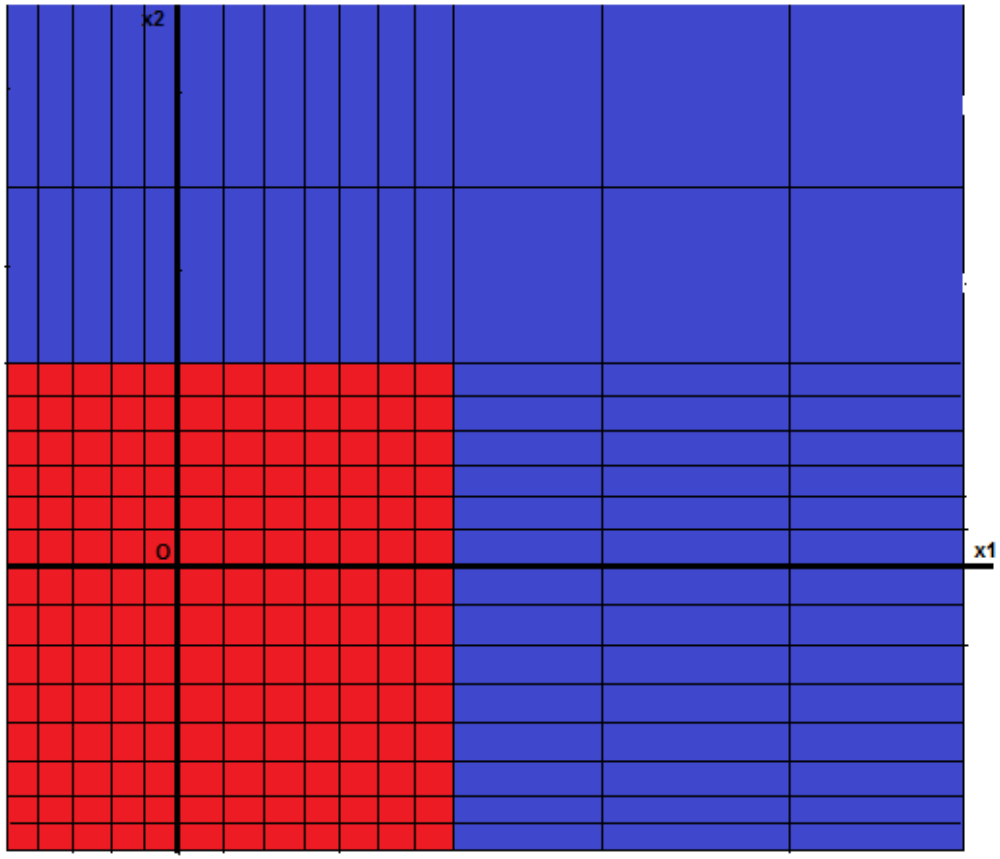
  - Discount factor: $\gamma = 1.0$
  - Learning rate: $\alpha = 0.7$

Design state variables

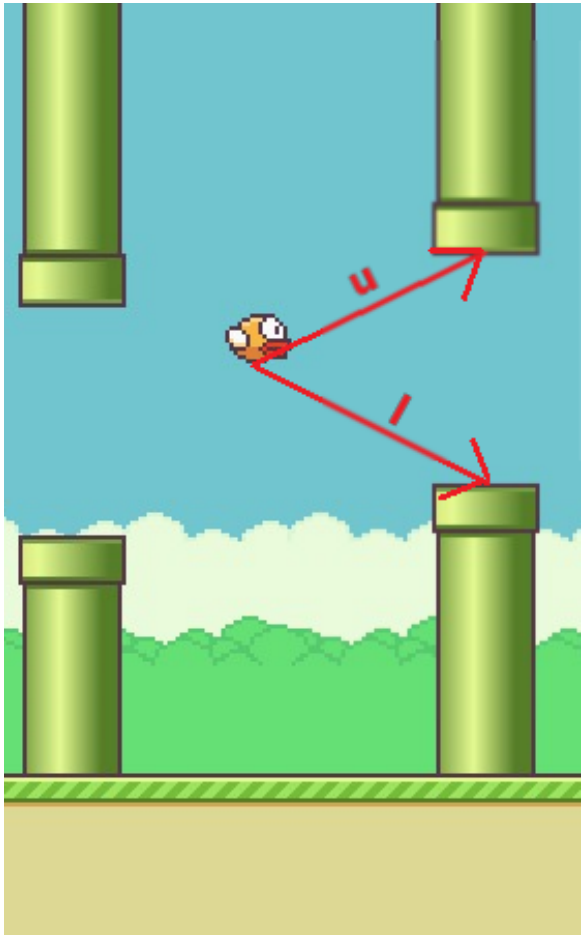


Discretization

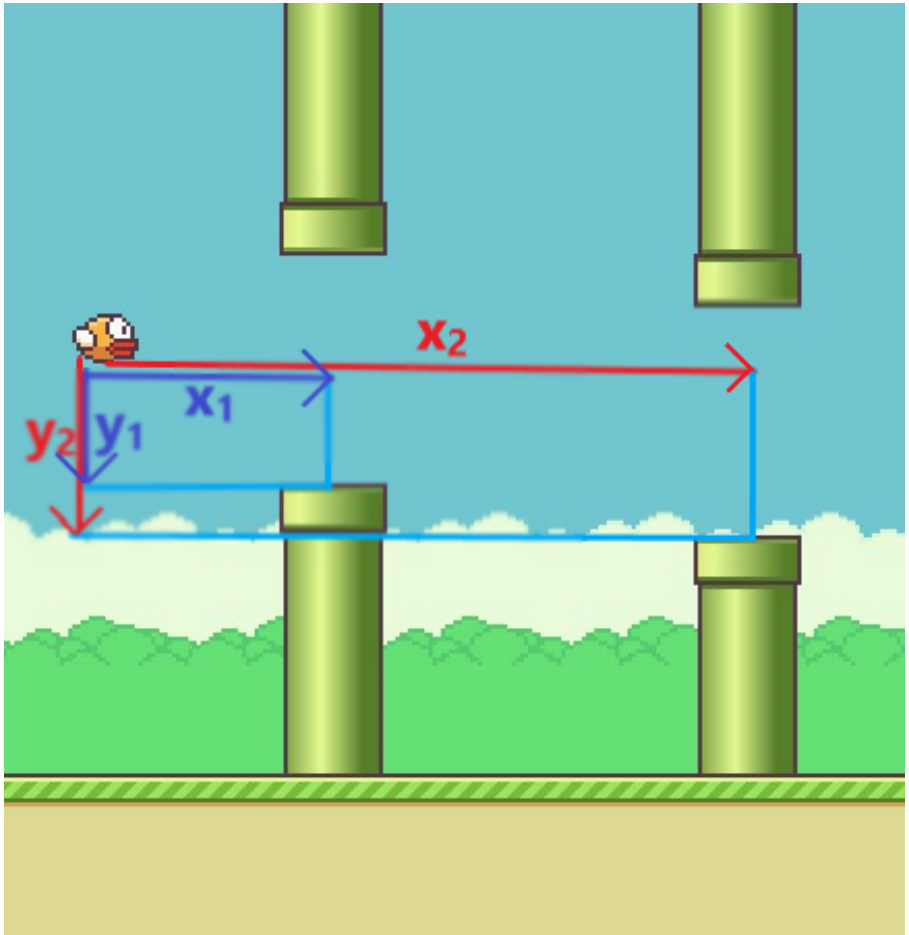
Discretizing states:



Design state variables

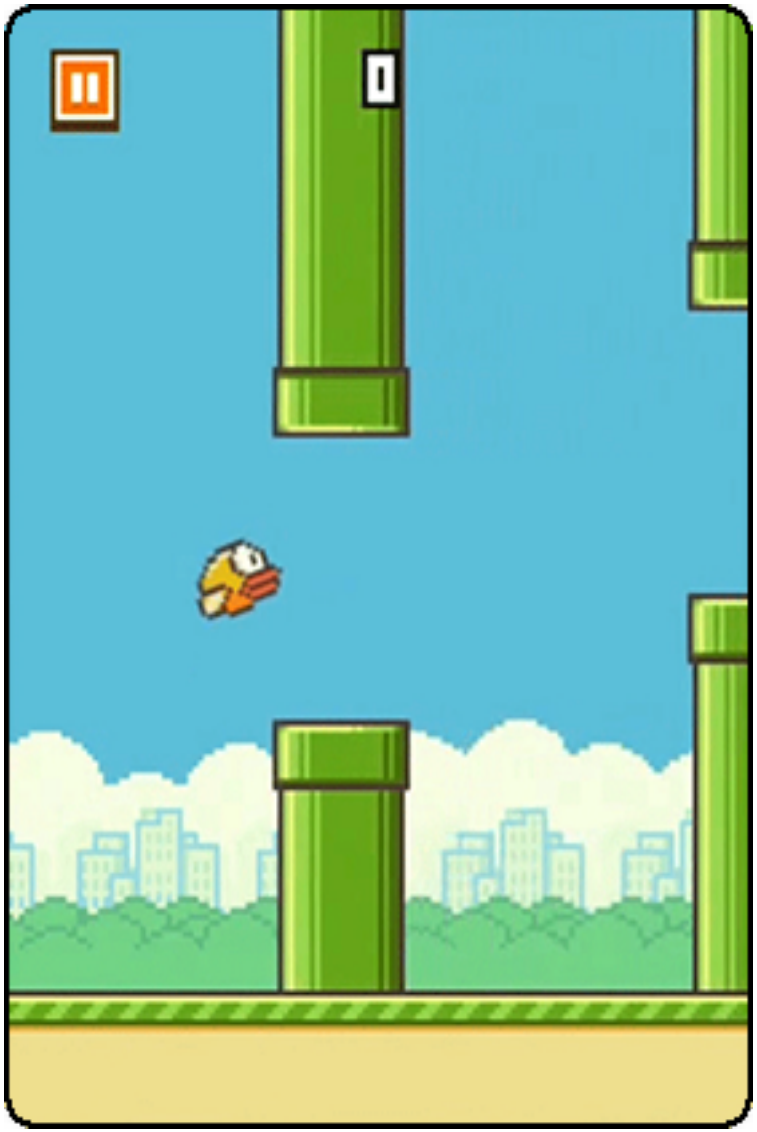


Design state variables

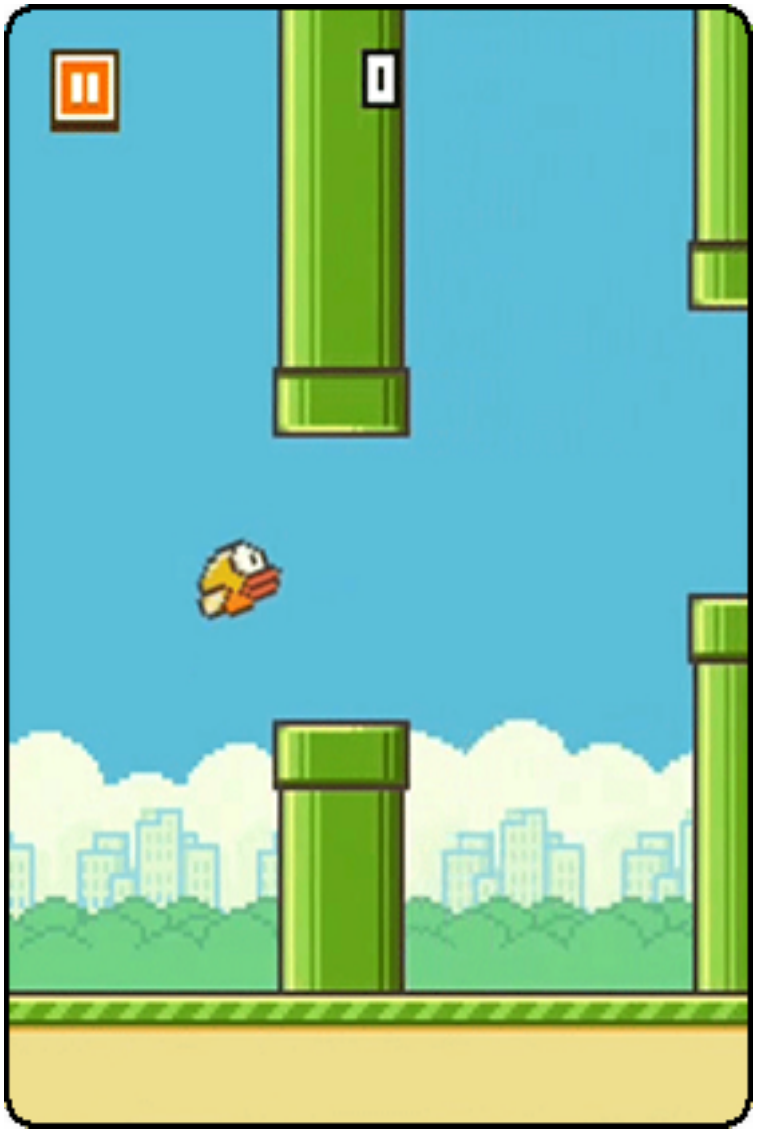


Training results

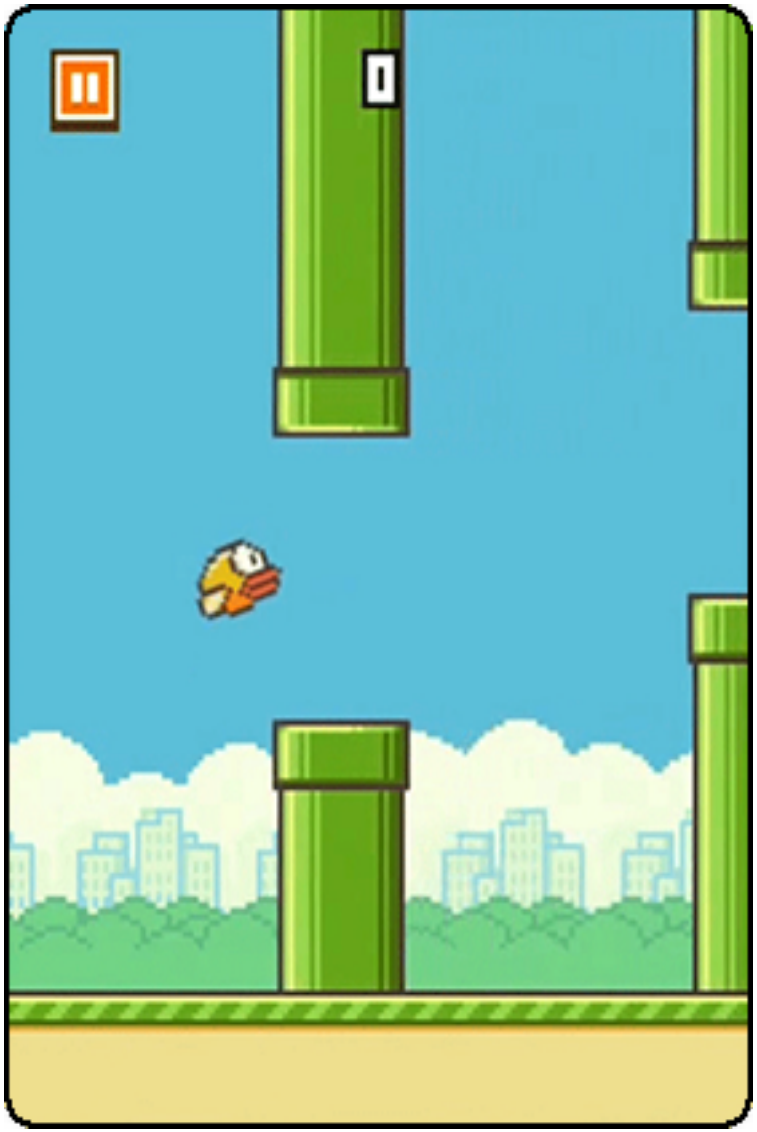
Beginner



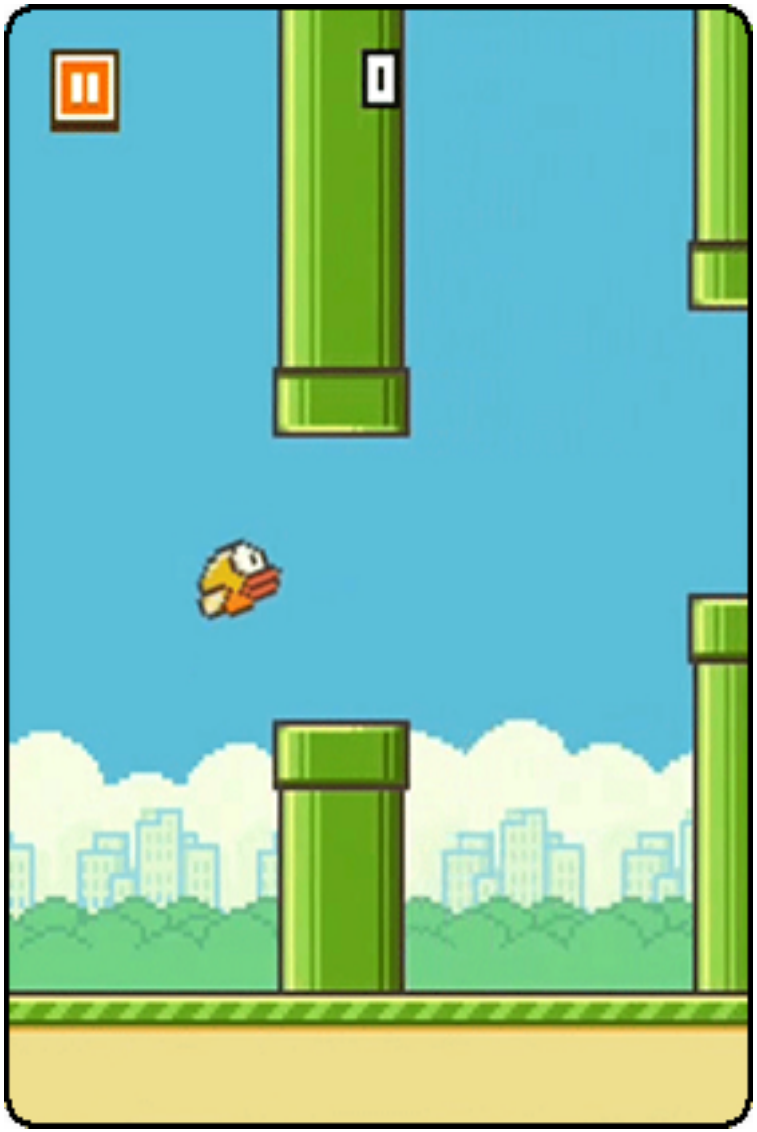
Amateur



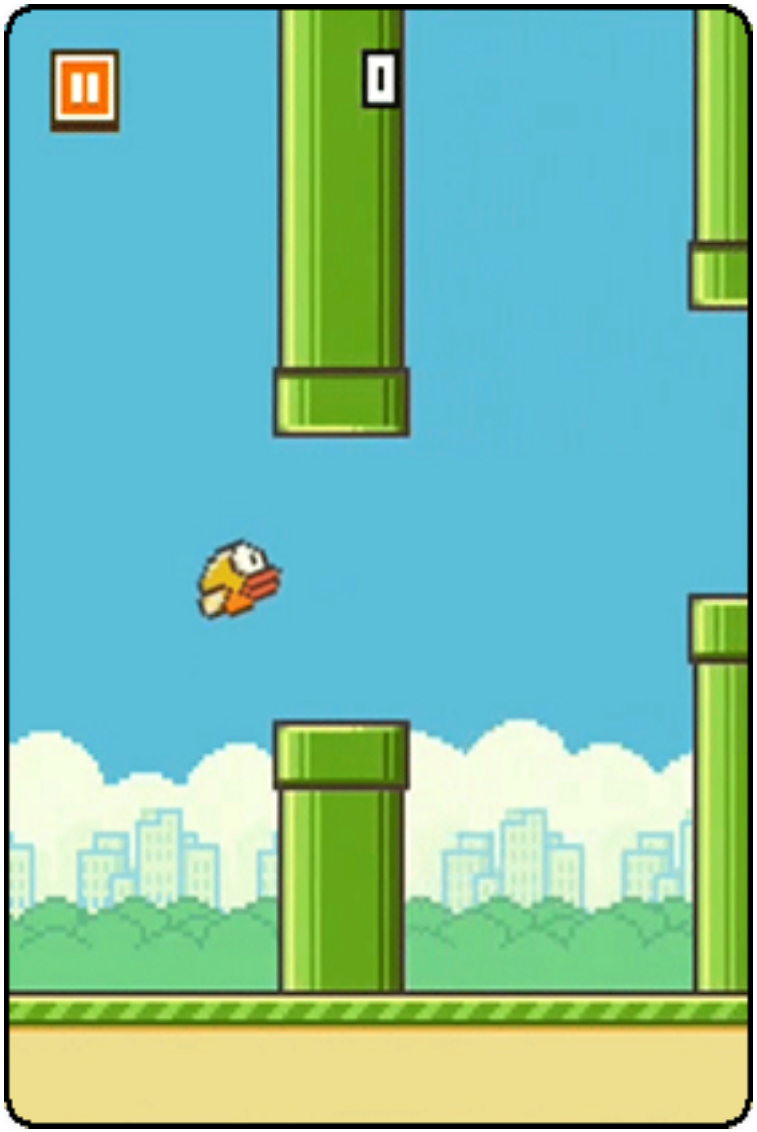
Intermediate



Pro

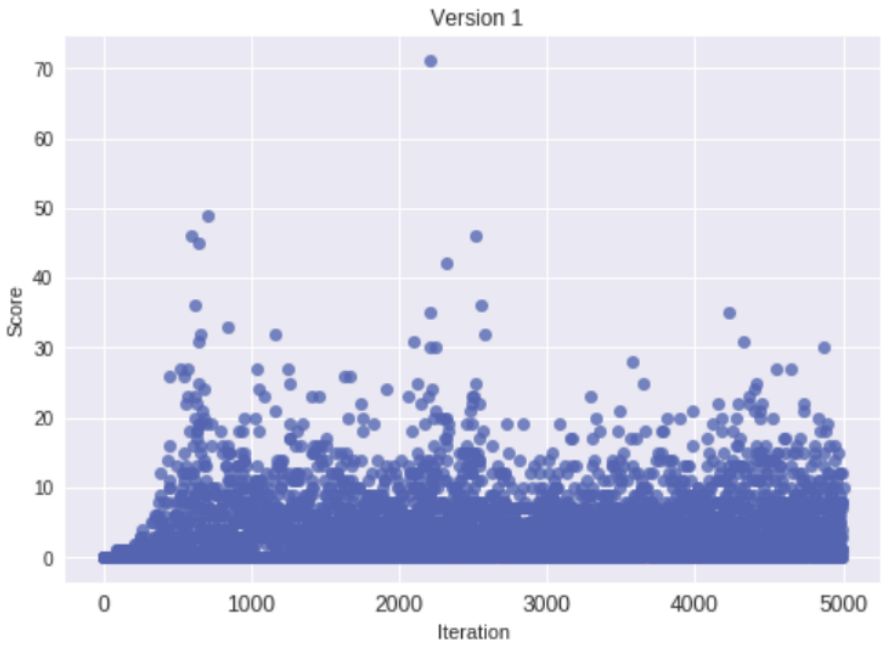


Speedup

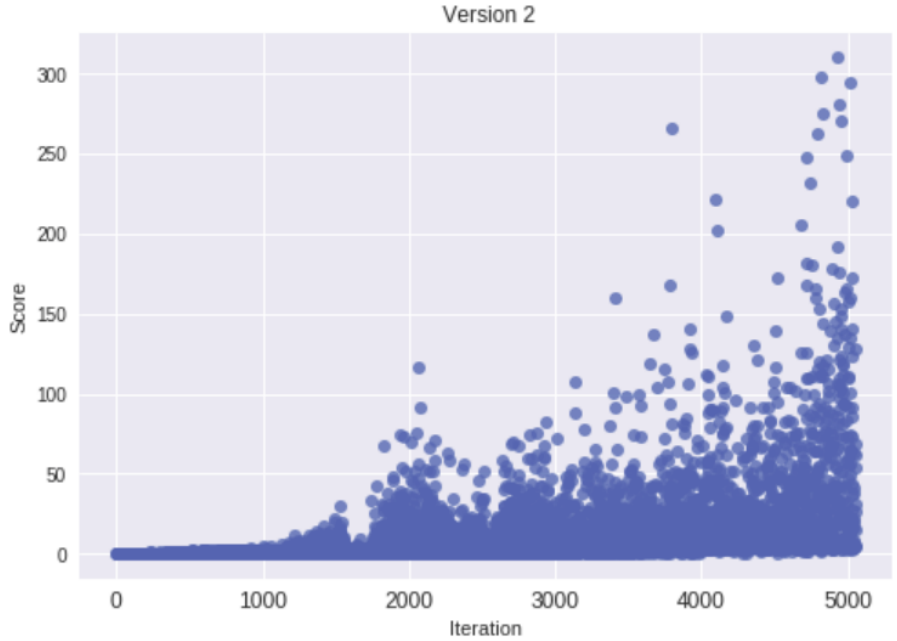


Plotting the results

Version 1: fine-grid $= 10 \times 10$.



Version 2: fine-grid $= 5 \times 5$.

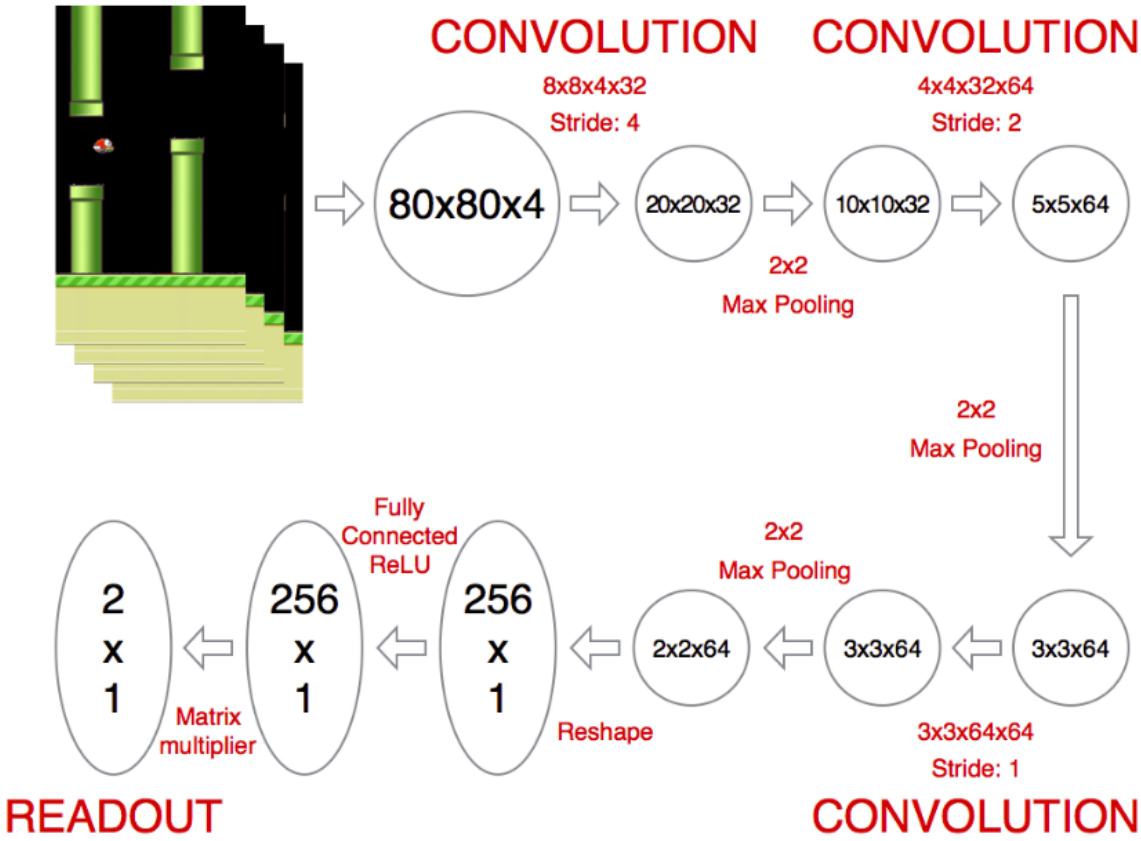


Learning state variables

Sử dụng Convolutional Neural Network để trích xuất các vector cơ sở.

Learning state variables

Sử dụng Convolutional Neural Network để trích xuất các vector cơ sở.



Conclusion

1. Hiểu được mối quan hệ giữa science (math) và engineering (coding) trong một Machine Learning project.

Conclusion

1. Hiểu được mối quan hệ giữa science (math) và engineering (coding) trong một Machine Learning project.
2. Reinforcement Learning: Từ designed basis (discretization) đến learned basis (Deep RL).

Conclusion

1. Hiểu được mối quan hệ giữa science (math) và engineering (coding) trong một Machine Learning project.
2. Reinforcement Learning: Từ designed basis (discretization) đến learned basis (Deep RL).
3. Trải nghiệm làm việc với game engine.

Source code for reference

- Flappy Bird Q-learning: https://github.com/chncyhn/flappybird-qlearning-bot
- Deep Learning Flappy Bird: https://github.com/yenchenlin/DeepLearningFlappyBird
- AlphaZero algorithm for Gomoku: https://github.com/junxiaosong/AlphaZero_Gomoku
- AlphaOne project and other projects of MaSSP 2018 Computer Science: https://github.com/masspvn/MaSSP2018

Thank you!

Xin chân thành cảm ơn quý vị khách mời đã chú ý lắng nghe!

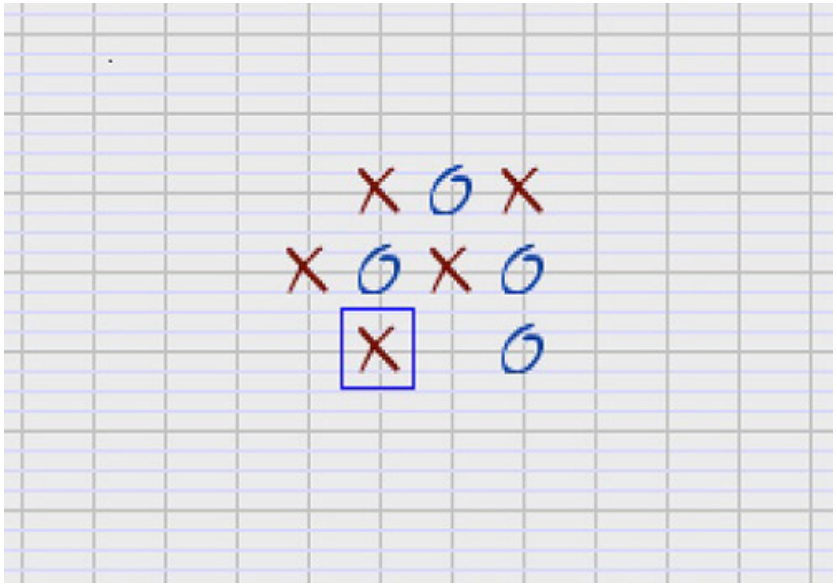
Bài toán Reinforcement Learning

Reinforcement Learning Problem:

- Task: States (e.g.: Environment images) $\xrightarrow{f^*}$ Optimal actions
- Performance: Total (discounted) rewards (Value function)
- Experience: trajectories of self-generated experience (s, a, s', r)
- Function Space: Linear hoặc nonlinear (deep) predictors
- Search Algorithm: Gradient-based (back-propagation)

Extension

Áp dụng thuật toán Reinforcement Learning để huấn luyện máy tính chơi một số game khác như: Gomoku, Mario...



Discussion

AlphaGO, AlphaZero: Learning + Planning