Исследование возможностей адаптации агентов обучения с подкреплением на примере игры Asteroids

Свидченко Олег Анатольевич

научный руководитель: д.н. в обл. ML Дэниел Куденко

СПб АУ НОЦНТ РАН

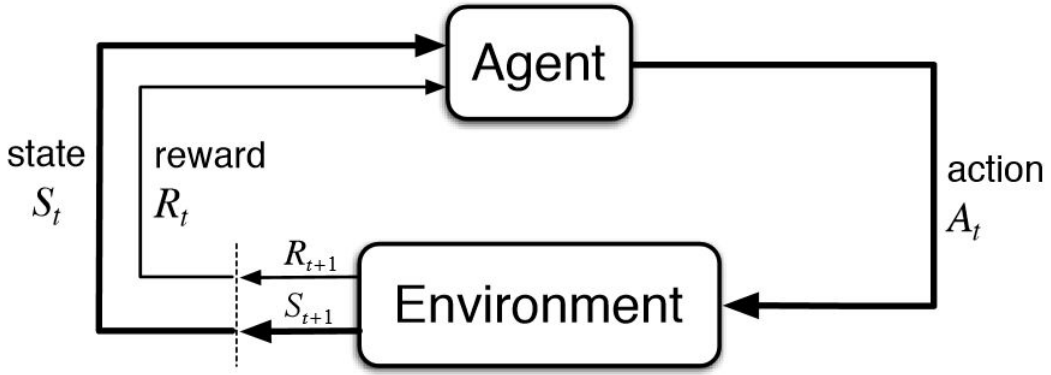
10.02.2018

Введение

- Для повышения эффективности компьютерных систем необходимо научиться их автоматизировать
- Большинство систем так или иначе взаимодействуют с человеком и другими системами
- Для более общего решения задачи необходима возможность адаптации к пользователю
- Скорость и эффективность адаптации важны

Введение

Модель обучения с подкреплением:



- Обучается, взаимодействуя со средой
- Реагирует на изменения в среде

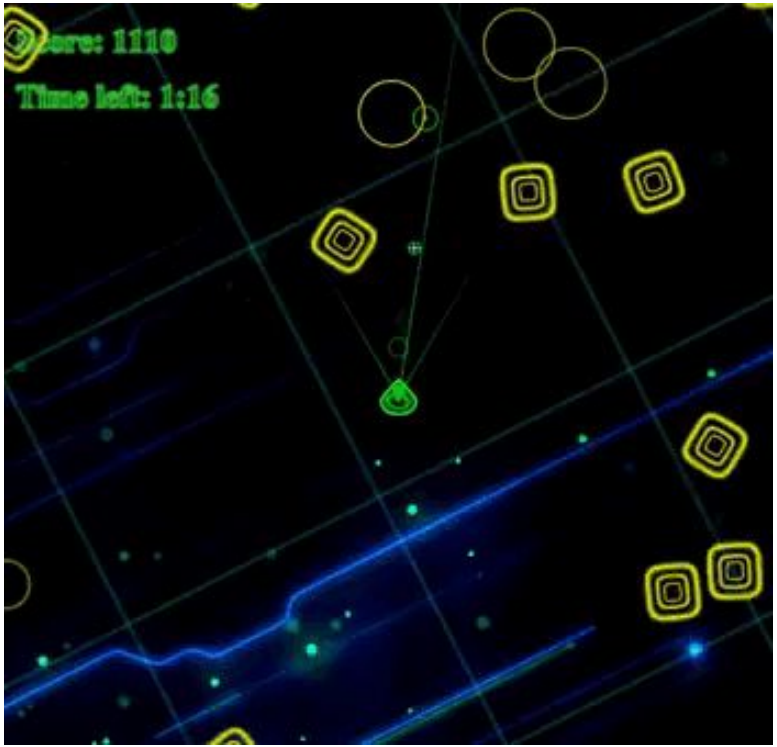
Что уже есть?

- Human-Robot Interaction: A historical perspective and current research trends
  - Michael Goodrich, Alan Schultz, Lanny Lin
  - Live document
  - https://goo.gl/7higwB

Задачи

- Реализовать среду для проведения исследования
  - Роли пользователей заранее зафиксированы
  - Оптимальность стратегии поведения одного пользователя зависит от стратегий других
  - Возможность при необходимости изменять сложность
- Исследовать возможности адаптации агентов RL к простому поведению
- Исследовать возможности адаптации агентов RL к более сложному поведению
- Провести эксперимент с участием реальных людей
  - Узнать, насколько отличается эффективность агентов при взаимодействии с людьми и при взаимодействии с автоматизированными пользователями

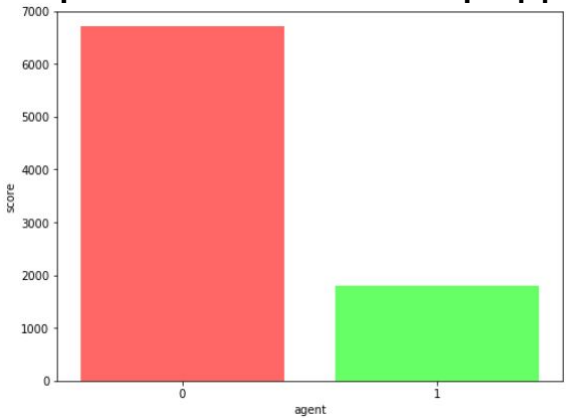
Игра Asteroids



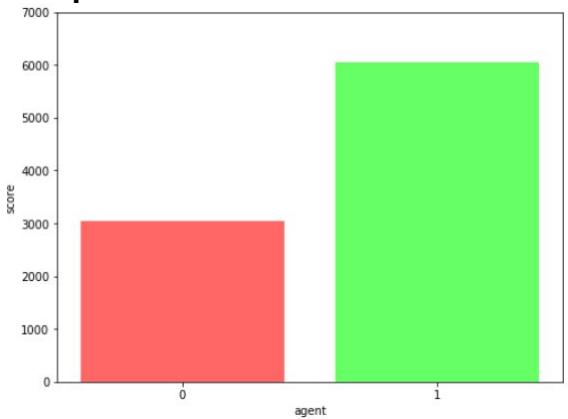
- Два игрока: стрелок и пилот
- Направление стрельбы и направление полета независимы
- Очень просто изменить сложность
- Нет ограничения по количеству жизней. Ограничение по времени

Игра Asteroids

Корабль летит вперед



Корабль стоит на месте

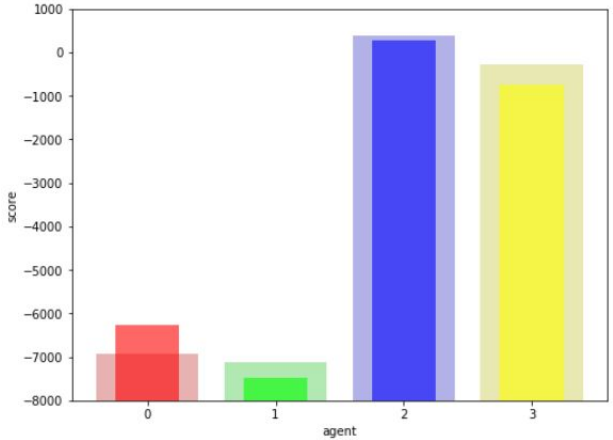


- Зеленая функция – счет агента, стреляющего в ближайший астероид
- Красная функция – счет агента, стреляющего в ближайшей по направлению полета астероид

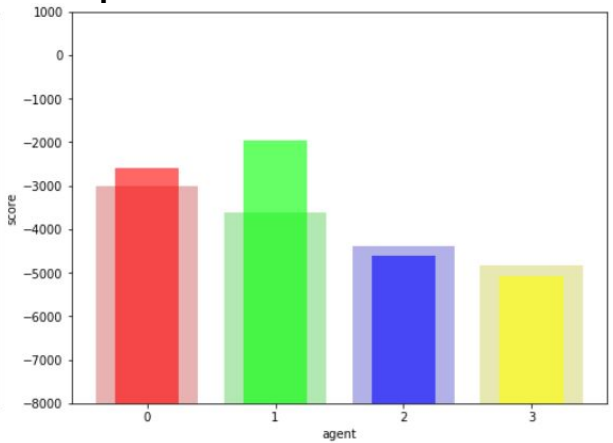
Проверим повторимость результата с участием агентов QLearning

QLearning и простые пилоты

Корабль летит вперед



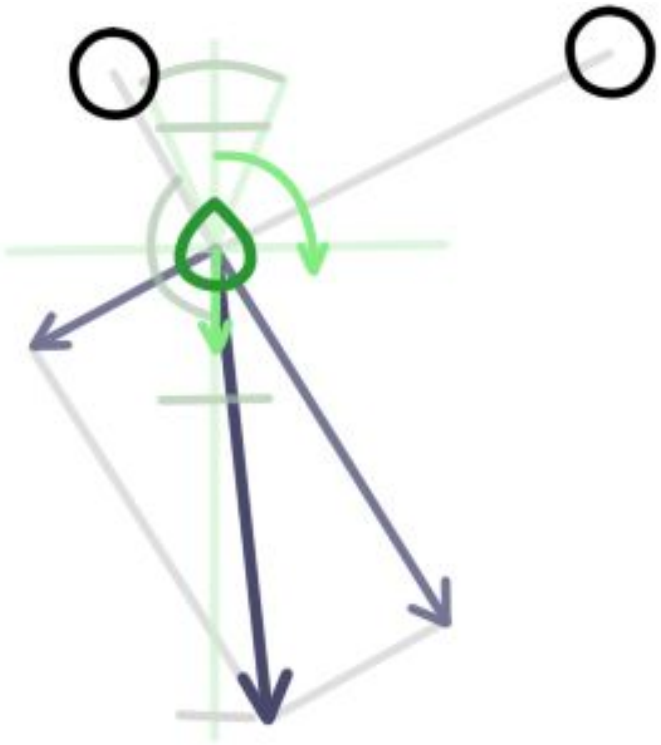
Корабль стоит на месте



- Красная и зеленая функции - агенты, знающие о ближайшем астероиде
- Желтая и синяя функции - агенты, знающие о ближайшем астероиде по направлению движения

Вывод: для агентов с соответствующим представлением мира результат повторяется.

Сложные пилоты

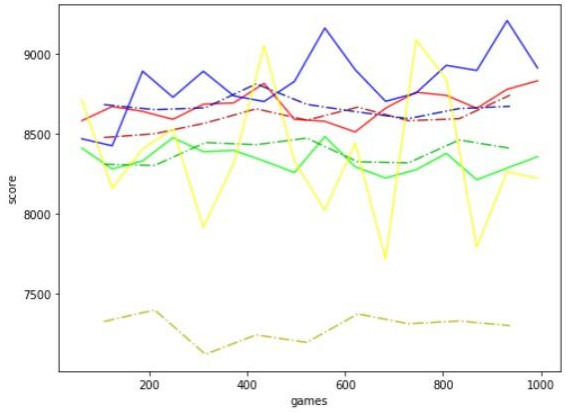


Реальные пользователи ведут себя сложнее, поэтому пилотов нужно сделать умнее.

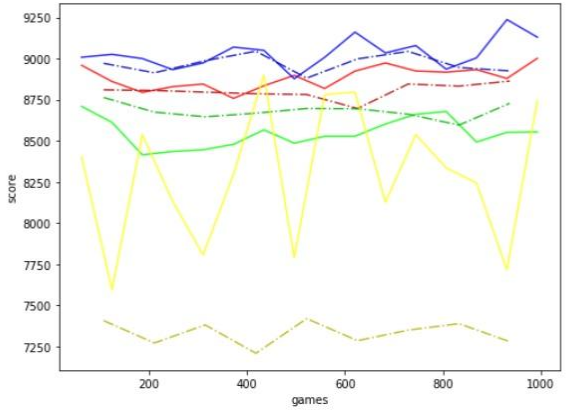
- Поворачиваем, если угол не попал в определенный сектор.
- Летим вперед или назад, если “сила” действует в соответствующем направлении.

QLearning и сложные пилоты

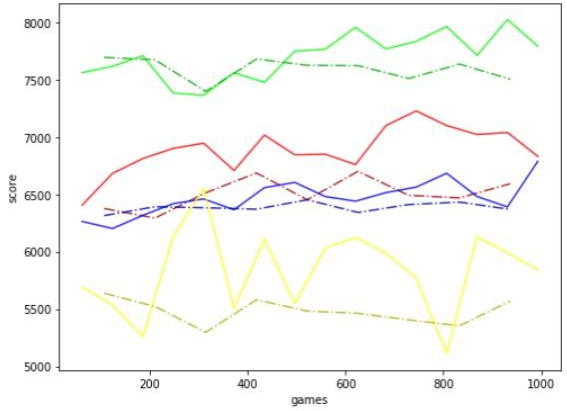
Fly forward



Fly away from asteroids



Avoid impact



- Теперь пилоты уклоняются от астероидов.
- Стрелки те же, что и при игре с простыми пилотами.

Большинство агентов при дообучении с конкретным пилотом почти не улучшают результаты. Следовательно, не адаптируются.

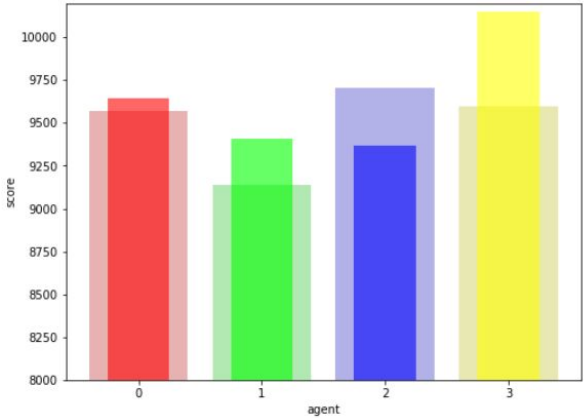
Линейное приближение

Улучшаем результат:

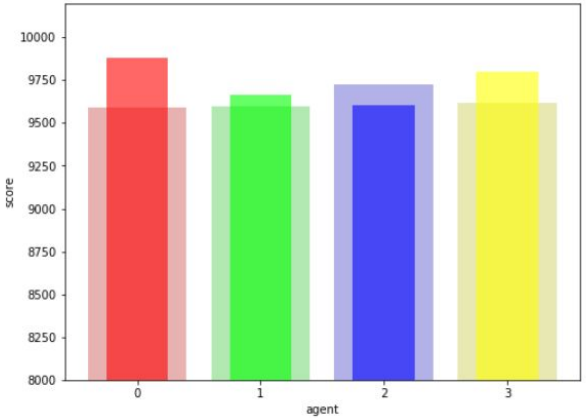
- Теперь выбираем стратегию на промежуток времени
- Считаем, что функция из стратегии в максимальный счет за игру хорошо приближается линейной функцией
  - Зависит от “силы”, действующей на корабль
  - Зависит от “силы”, действующей на орудие
  - Зависит от угла между кораблем и орудием

Линейное приближение

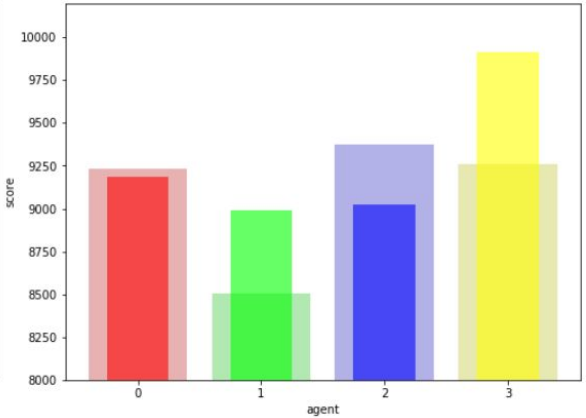
Fly forward



Fly away from asteroids

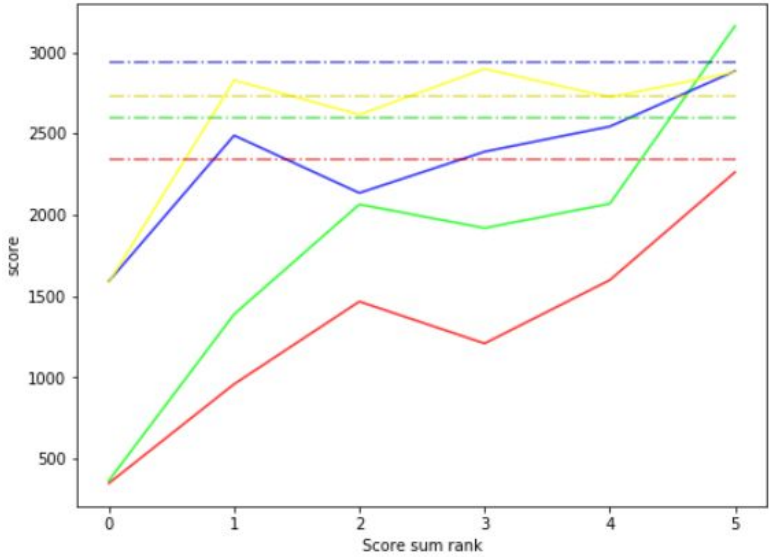


Avoid impact



- Разные наборы входных данных приводят к разным результатам и к разной устойчивости агентов
- При правильном наборе входных данных агенты способны значительно улучшать результат при дообучении

Эксперимент с реальными людьми



- Красная и зеленая функция – агенты QLearning.
- Синяя функция – линейное приближение
- Желтая функция – простой стрелок, стреляющий в ближайший астероид

Пунктиром отмечен средний счет, набираемый агентами соответствующих цветов при игре со сложными пилотами.

Выводы

- Агенты QLearning плохо адаптируются к сложному поведению пилота
- Агенты, основанные на методе линейного приближения, лучше адаптируются к сложному поведению пилота, однако нужно правильно выбирать входные данные
  - Можно ли еще как-то улучшить агентов?
- Агенты, обученные на автоматизированных пилотах, значительно хуже играют с реальными пользователями
  - В чем разница между реальным пилотом и автоматизированным?
  - Как зависит эффективность агентов от навыка игры пилота?
  - Адаптируются ли пилоты к агентам? Есть ли зависимость между навыком игры и эффективностью адаптации?