Применение структурного машинного обучения для игр с неполной информацией
Обзор работ по теме

Мисник А.С.
misnik.as@phystech.edu

Центр когнитивного моделирования МФТИ
ФИЦ ИУ РАН

2024 – Научно-исследовательский семинар

Введение

- Большинство моделей машинного обучения хранят переработанную информацию об объектах и результаты обучения отдельно от их исходных признаков
- Единственный общепринятый метод передачи знаний между моделями - дистилляция

Мотивация

- Получение структурного представления эталонных политик в обучении с подкреплением
- Обобщение полученного результата - формирование набора человекочитаемых правил (например, для ручного анализа экспертами в предметной области)

Постановка задачи

Выберем набор бинарных признаков исходных состояний и зафиксируем отображение F $(\forall s \in S : F(s) = \{f_1, f_2, .., f_n\}, f_i \in \{0, 1\})$

- Целью агента в классическом обучении с подкреплением является выбор оптимальной политики $A_t \sim \pi(\cdot|S_t)$
- Цель агента в модели задачи - порождение множества гипотез H $(\forall h_i \in H : h_i = \{F_i, A_i\}, F_i \subset \{0, 1\}^n)$, где каждая гипотеза имеет свойство $\forall s_t \in S : F(s_t) \in F_h \Rightarrow A_h \sim \pi(\cdot|s_t)$

Обзор ВКФ

- Для выбора оптимальной стратегии используется Monte Carlo Tree Search (MCTS) на деревьях игр
- Между состояниями вводится функция сходства $sim(S_1, S_2) = F(S_1) \ \& \ F(S_2)$
- С использованием операции сходства формируются гипотезы о выборе действий в похожих состояниях

Алгоритм ВКФ

Шаг алгоритма:

1. Породить гипотезу-кандидата с использованием цепи Маркова
2. Для каждого из состояний, затрагиваемых в гипотезе, проверить оптимальность действия (используя результат симуляции MCTS для этого состояния)
3. Подтвердить гипотезу, если действие оптимально для всех затрагиваемых состояний

Процесс алгоритма:

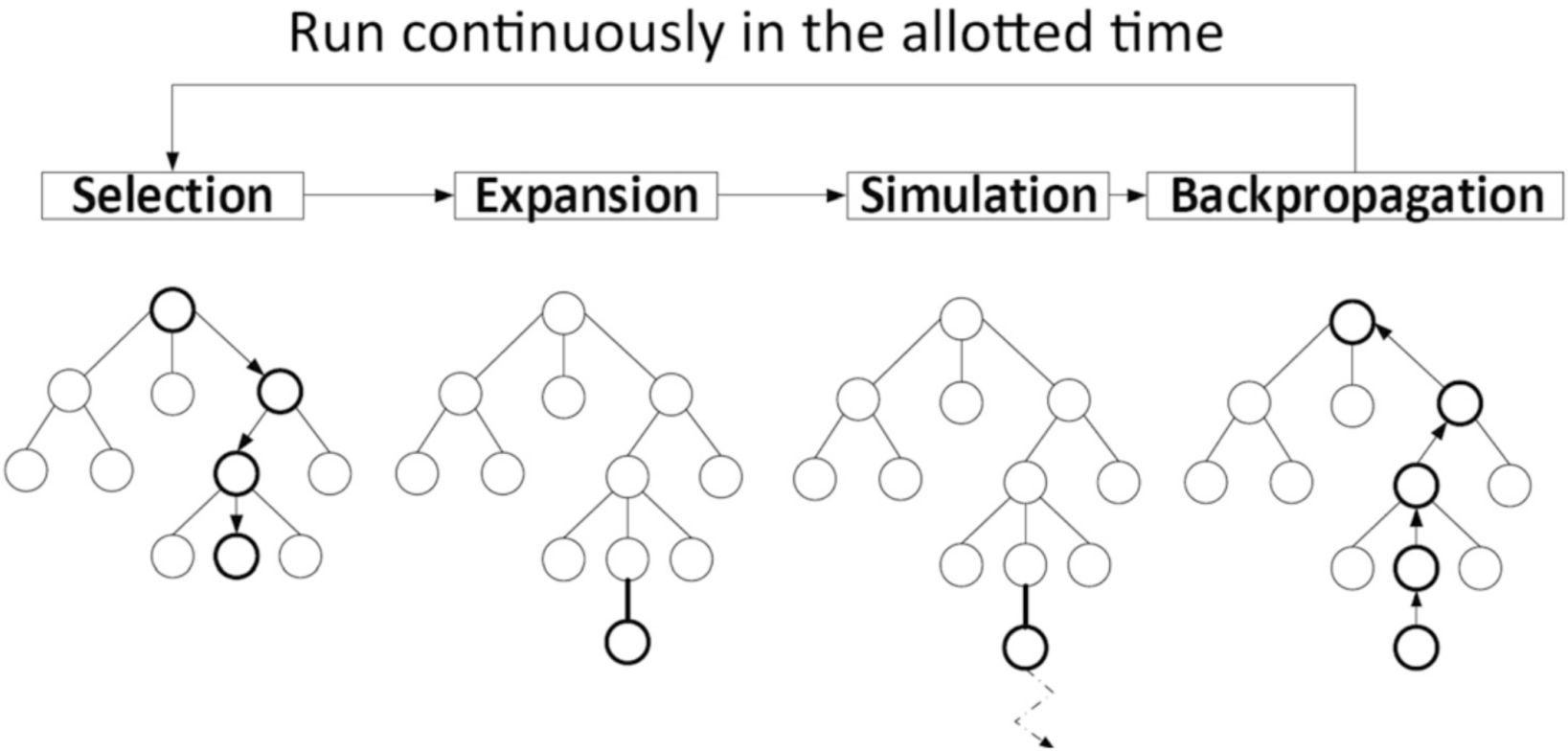
1. Породить N подтвержденных гипотез, используя шаги алгоритма
2. Вычислить эмпирический риск применения порожденных гипотез
3. Повторить шаги 1 и 2 с увеличенными N, пока эмпирический риск не перестанет уменьшаться

Monte-Carlo Tree Search

Алгоритм для сбора статистики и принятия решений в комбинаторных пространствах, описываемых Марковским процессом принятия решений.

- Был предложен в 2006 году в контексте создания ИИ для игры в Го
- В 2016 году была представлена программа AlphaGo, основанная на MCTS
- Применяется в задачах планирования, контроля и комбинаторной оптимизации

Схематическое описание MCTS



Шаги алгоритма Monte-Carlo Tree Search

Алгоритм MCTS

В рамках заданных ресурсов повторяется 4 шага алгоритма:

1. **Selection**
   Перемещение по дереву игры, начиная с корневого узла
2. **Expansion**
   Выбор ещё не исследованного дочернего узла и расширение дерева
3. **Simulation**
   Симуляция игры до терминального состояния для узла, добавленного на предыдущем шаге
4. **Backpropagation**
   Обновление информации на пути от исследуемого узла до корня на основе результатов симуляции

Применение в играх с неполной информацией

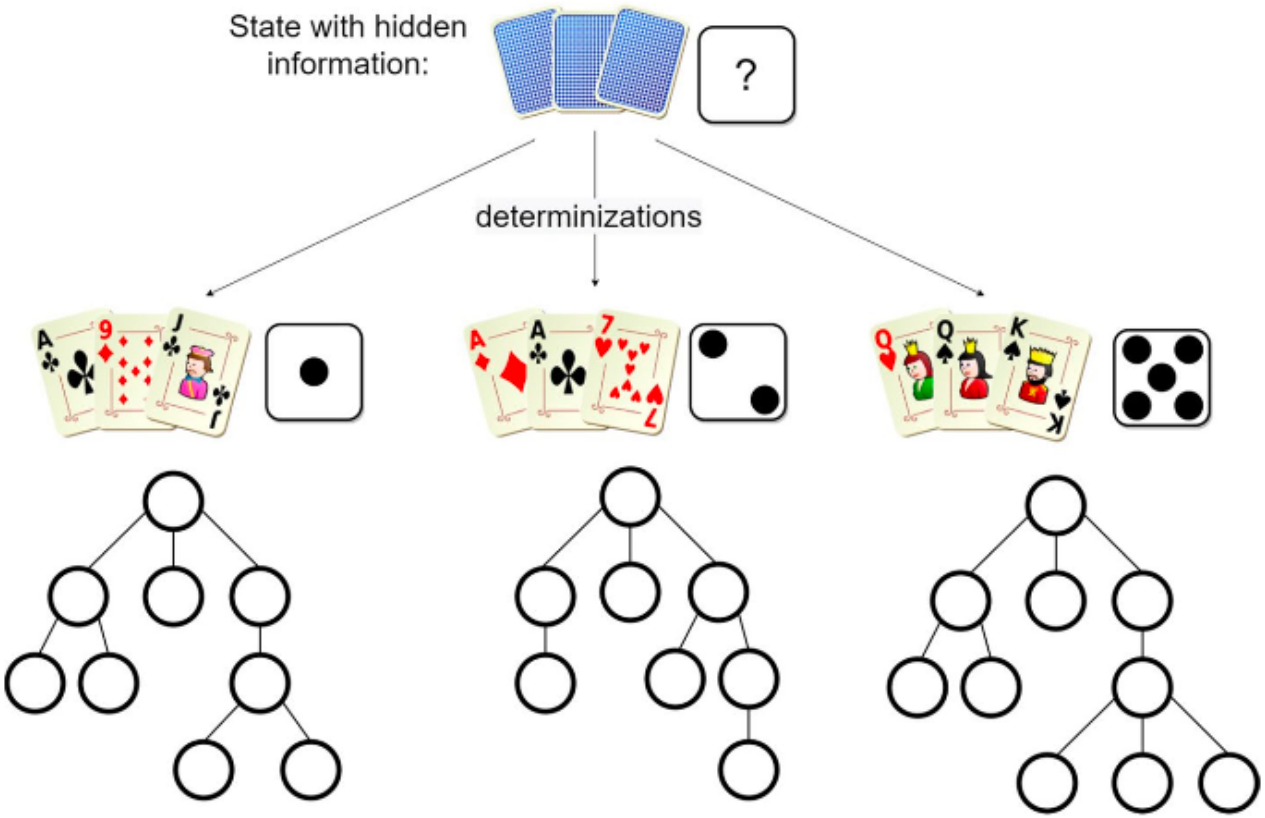
Неполнота инфромации о состояниях увеличивает ветвление, поскольку каждый скрытый элемент порождает множество состояний игры с распределением вероятностей на нем.

Способы решения:

- Детерминация состояний
- Информационные наборы
- Детерминация переходов

Детерминация состояний

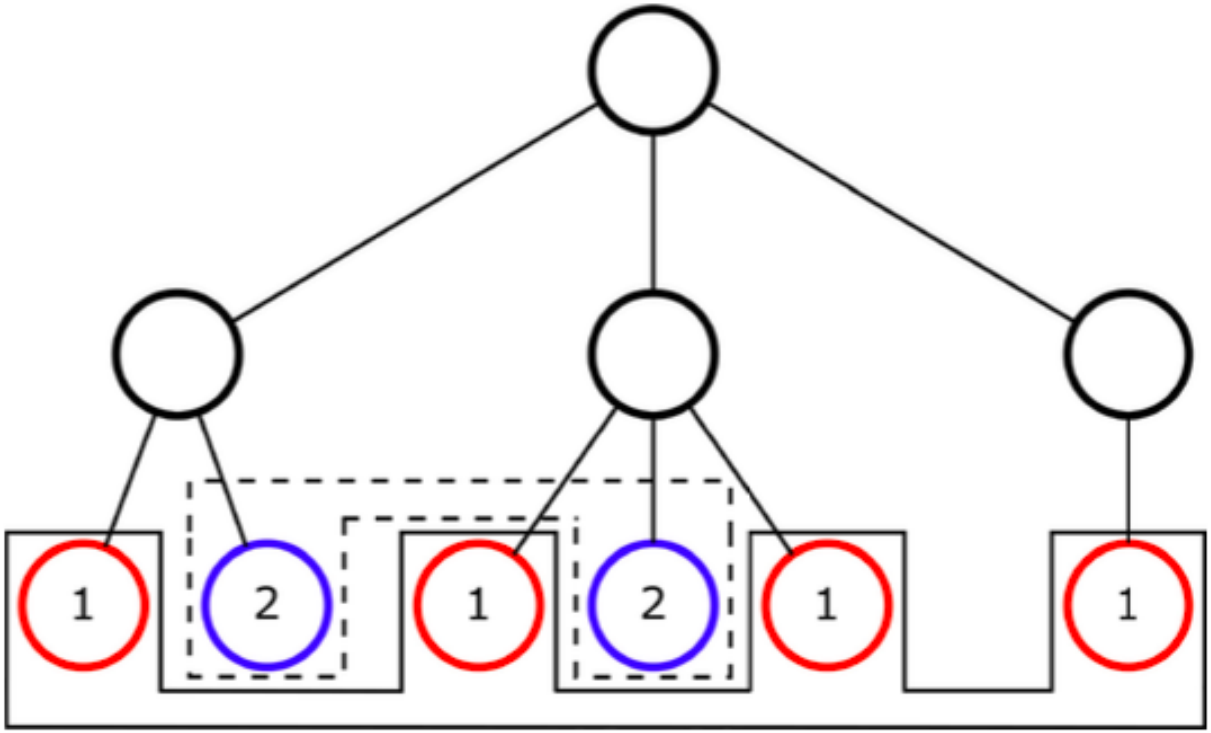
Формирование игр с полной информацией из пространства возможных состояний.



Пример детерминации состояний

Информационные наборы

Выделение наборов состояний, неотличимых с точки зрения агента; позволяет решить проблему "раздувания" после детерминации состояний.



Пример объединения - состояния объединены в 2 группы

Детерминация переходов

Замена случайных переходов на детерминированные путем ограничения правил и обучение отдельных агентов в условиях детерминированности.

Short name	Description	Trees	Simulations per tree	UCT constant	Win/loss reward	Reward discount	Move pruning	Simulation strategy	Tree structure
AP	All Possible Deals / Uniform Random Rollouts	40	250	1.5	1 / 0	1	None	Uniform Random	All Possible Deals
BA	Baseline	40	250	1.5	1 / 0	0.99	Land	Reduced Rules	Unlimited Degree
IL	Interesting Simulations (Leaf. 1% of sim budget)	40	250	1.5	1 / 0	0.99	Land	**Reduced Rules / Interesting (Leaf)**	Unlimited Degree
IR	Interesting Simulations (Root. 5% of sim budget)	40	250	1.5	1 / 0	0.99	Land	**Reduced Rules / Interesting (Root)**	Unlimited Degree
NL	Negative Reward for Loss	40	250	1.5	**1 / -1**	0.99	Land	Reduced Rules	Unlimited Degree
MP	Dominated Move Pruning	40	250	1.5	1 / 0	0.99	**Dominated**	Reduced Rules	Unlimited Degree
BT	Binary Tree (Descending Mana Cost)	40	250	1.5	1 / 0	0.99	Land	Reduced Rules	**Binary**

Пример создания ансамбля моделей с ограниченными правилами (для игры Magic: The Gathering)

Выводы

- Monte Carlo Tree Search - гибкий инструмент для решения задач последовательного принятия решений, популярность которого выросла в последние несколько лет
- Существует несколько способов оптимизации MCTS для игр с неполной информацией
- Большинство методов, использующих MCTS в обучении с подкреплением, создают необъяснимые на человеческом языке политики

Спасибо за внимание!

misnik.as@phystech.edu