Задача о разборчивой невесте

Тунёв Иван, Б05-875 группа

План

- Предыстория и постановка задачи;
- Оптимальная стратегия;
- Вывод оптимальной стратегии: немного про управляемые марковские процессы и сопутствующие уравнения;
- Какие можно составить ещё стратегии в задаче;
- Численное сравнение;

Предыстория и постановка задачи;

- Сформулировал Мартин Гарднер в 1950-е годы;
- Решили: Е. Дынкин (частный случай, 1963 г.), С. Гусейн-Заде (обобщение, 1966 г.);
- Английские названия: Secretary, Dating, Sultan`s dowry, best choice problem;
- Невеста ищет себе мужа среди n женихов;
- Если жених не понравился невесте, то он выбывает из игры, а невеста идёт смотреть следующего;
- Если жених понравился, то процесс останавливается (естественно?);
- **Женихи идут в случайном порядке! Мы не знаем, кто когда будет;**

Оптимальная стратегия

- Оптимальная стратегия -- пропустить первую 1/e часть, а потом выбрать первого лучшего среди оставшихся;
- Плохо: есть шанс остаться одной;
- Данная стратегия позволит выбрать наилучшего жениха с вероятностью 1/e (даже больше чем ⅓!);
- Вопрос: а как прийти к такому решению?
- Ответ даст теория управляемых марковских процессов и уравнение Вальда-Беллмана;

Про управляемые марковские цепи

- Марковская цепь с конечным множеством состояний S;
- На каждом шаге система находится в каком-то из этих состояний, мы этой системой управляем: в зависимости от выбранного действия a из множества A определяется вероятность перехода в другое состояние в следующий квант времени;
- В каждый момент времени получаем вознаграждение;
- Цель -- максимизировать итоговое вознаграждение;

Про управляемые марковские цепи

Немного обозначений:

- Вероятность перехода между соотв. состояниями при выполнении действия a $p(s, a; s')$
- Вознаграждение r(s, a) и его матожидание R(s, a)
- Функция цены (ожидаемый выигрыш при оптимальной стратегии при изначальном состоянии s) $V^*(s)$

Про управляемые марковские цепи

И теперь уже конкретно о нашей цели:

$$V^*(s) = \max_{a(\cdot)} \mathbb{E} \sum_{t=0}^{\infty} \gamma^t r(s_t, a(s_t)), \ \ s_0 = s.$$

γ -- коэффициент дисконтирования

Динамическое программирование?

- Функция цены удовлетворяет уравнению Вальда-Беллмана

$$V^*(s) = \max_{a \in A} \left(R(s,a) + \gamma \sum_{s' \in S} p(s,a;s') V^*(s') \right)$$

- Условие для нахождения оптимальной стратегии

$$a(s) = \arg\max_{a \in A} \left(R(s,a) + \gamma \sum_{s' \in S} p(s,a;s') V^*(s') \right).$$

Динамическое программирование

Идея доказательства (здесь $s = s_0, s_1 = s'$):

$$
\begin{aligned}
V^*(s) &= \max_{a(\cdot)} \mathbb{E} \sum_{t=0}^{\infty} \gamma^t r(s_t, a(s_t)) = \\
&= \max_{a(\cdot)} \mathbb{E} \left(r(s_0, a(s_0)) + \gamma \sum_{t=0}^{\infty} \gamma^t r(s_{t+1}, a(s_{t+1})) \right) = \\
&= \max_{a \in A} \left(R(s_0, a) + \gamma \mathbb{E}_{s_1} V^*(s_1) \right) = \\
&= \max_{a \in A} \left(R(s, a) + \gamma \sum_{s_1} p(s, a; s_1) V^*(s_1) \right).
\end{aligned}
$$

Непосредственно поиск стратегии

- Введём управляемую маркову цепь;
- $\gamma = 1,\ S = \{1, 2, \ldots, N, End\},\ N = 1000$
- $A = \{\text{не выбрать, выбрать}\}$
- Перейдём к расчёту соответствующих функций вознаграждения и переходных вероятностей:

$$R(s = End,\ a = \text{выбрать}) = 0,\ R(s,\ a = \text{не выбрать}) = 0;$$
$$R(s,\ a = \text{выбрать}) = s/N,\ s = 1, \ldots, N,$$

Непосредственно поиск стратегии

$$r(s, a = \text{выбрать}) = \begin{cases} 1, & \text{с вероятностью } s/N, \\ 0, & \text{с вероятностью } 1 - s/N. \end{cases}$$

С переходными вероятностями немного сложнее:

$$p(s,\ a = \text{выбрать};\ s') = 0,\ \ s' \neq End,$$
$$p(s,\ a = \text{выбрать};\ s' = End) = 1;$$
$$p(s = End,\ a,\ s' = End) = 1;$$
$$p(s,\ a = \text{не выбрать};\ s' = End) =$$
$$= \mathbb{P}\left(\begin{array}{c} s\text{-й претендент лучше всех, если} \\ \text{известно только, что он лучше предыдущих} \end{array}\right) = \frac{s}{N};$$

Непосредственно поиск стратегии

$$p(s,\, a = \text{не выбрать};\, s') =$$

$$= \mathbb{P}\begin{pmatrix} s'\text{-й претендент – первый кто, лучше } s\text{-го,} \\ \text{если известно, что } s\text{-й лучше предыдущих} \end{pmatrix} =$$

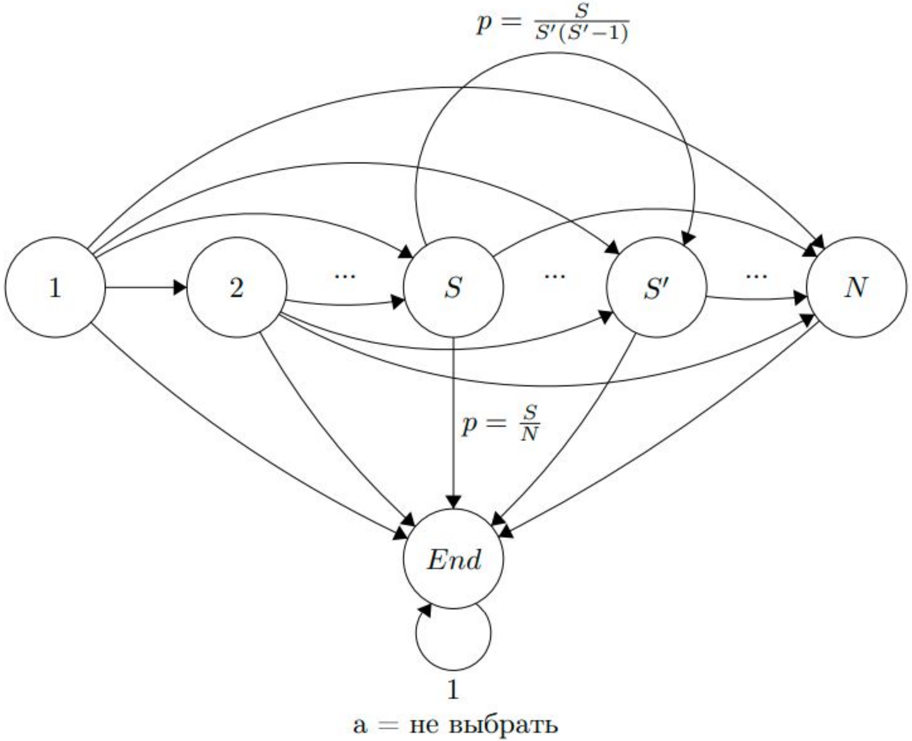
$$= \frac{\mathbb{P}\begin{pmatrix} s\text{-й претендент лучше предыдущих и} \\ s'\text{-й претендент – первый кто, лучше } s\text{-го} \end{pmatrix}}{\mathbb{P}(s\text{-й претендент лучше предыдущих})} = \frac{s}{s'(s'-1)},$$

поскольку

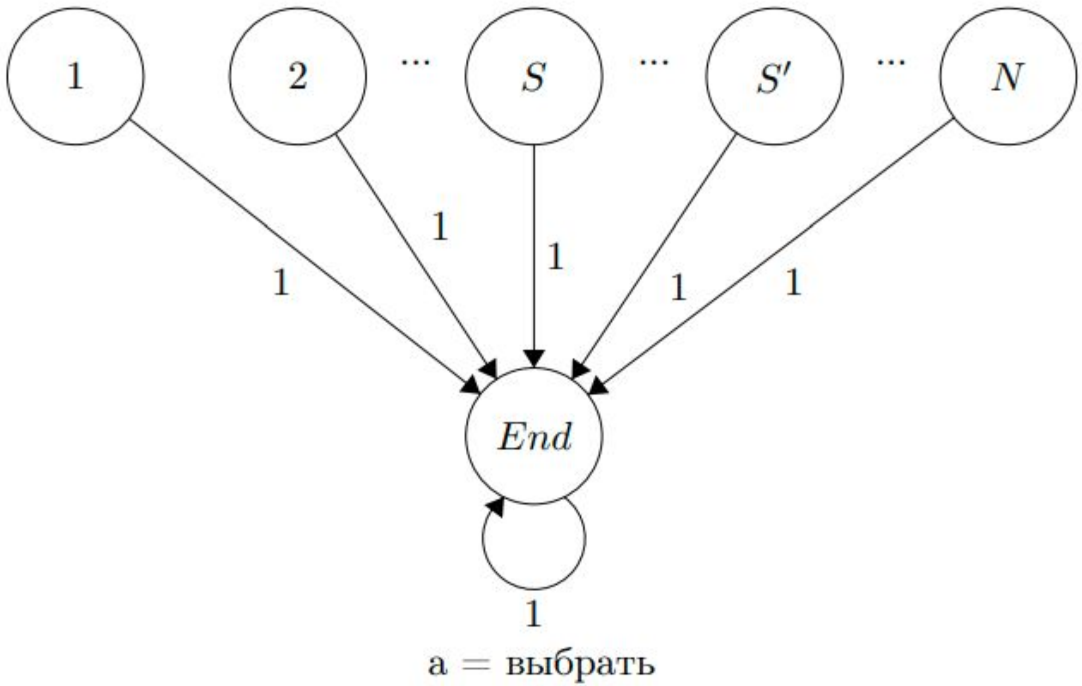
$$\mathbb{P}(s\text{-й претендент лучше предыдущих}) = \frac{(s-1)!}{s!} = \frac{1}{s},$$

$$\mathbb{P}\begin{pmatrix} s\text{-й претендент лучше предыдущих и} \\ s'\text{-й – первый кто, лучше } s\text{-го} \end{pmatrix} = \frac{(s'-2)!}{s'!} = \frac{1}{s'(s'-1)}.$$

Непосредственно поиск стратегии (граф)



Непосредственно поиск стратегии (граф)



Непосредственно поиск стратегии

- Выпишем Вальда-Беллмана:

$$V^*(s) = \max_{a \in A} \left(R(s,a) + \gamma \sum_{s' \in S} p(s,a;s') V^*(s') \right),$$

$$V^*(s) = \max \left(\frac{s}{N}, \sum_{s'=s+1}^{N} \frac{s}{s'(s'-1)} V^*(s') \right), \quad s = 1, \ldots, N{-}1; \; V^*(N) = 1.$$

- Если максимум достигнут на первом аргументе, то a(s) -- это выбрать, если на втором -- то не выбрать. Здесь всё разрешимо, покажем это:

Непосредственно поиск стратегии

Для этого определим $s^*(N)$ из уравнения

$$\frac{1}{s^*} + \frac{1}{s^*+1} + \cdots + \frac{1}{N-1} \leq 1 \leq \frac{1}{s^*-1} + \frac{1}{s^*} + \cdots + \frac{1}{N-1}.$$

Можно показать, что

$$s^*(N) \simeq \left[\frac{N}{e}\right].$$

Введем

$$V^* = \frac{s^*(N)-1}{N}\left(\frac{1}{s^*(N)-1} + \frac{1}{s^*(N)} + \cdots + \frac{1}{N-1}\right) \simeq \frac{1}{e}.$$

Тогда

$$V^*(s) = \begin{cases} V^*, 1 \leq s \leq s^*(N), \\ s/N, s \geq s^*(N). \end{cases}$$

Другие стратегии и обобщения

- Невеста всегда выбирает k по очереди жениха (очевидно, что провал, т.к. вероятность поймать лучшего очень мала);
- Можно положить, что нам нужен не обязательно лучший, но можно и просто хорошего жениха. Тогда после отсева первых 1/e выбирать не первого лучшего, а первого, кто отстоит от лучшего по тем или иным показателям не более чем на фиксированное расстояние;
- Последнее похоже на обобщение задачи, когда невеста выбирает одного из m лучших женихов;

Численное моделирование и анализ

- Рассмотренные стратегии:
- Невеста играет по эталонным правилам;
- Невеста выбирает случайного жениха;
- С вероятностью s/n берём s по порядку жениха;
- Берём первого, вес которого не больше чем в 0.75 хуже чем у лучшего;
- Rate: доля попаданий в лучшего;

Численное моделирование и анализ. Итоги

- Выбор случайного кандидата даёт околонулевые результаты (как и ожидалось);
- Внезапно (?) обнаружилась стратегия, давшая на рассмотренных значениях n результаты сильно лучшие чем у оптимального;
- Оптимальный алгоритм стартовал с высокими характеристиками (под 60%), но довольно быстро сошёлся к 37%;

Литература и источники