Онлайн ценообразование с помощью модели структурированных многоруких бандитов

Гунаев Руслан Гуламович

Московский физико-технический институт
Факультет управления и прикладной математики
Кафедра интеллектуальных систем

Научный руководитель Ю. Дорн

15 декабря 2020 г.

Онлайн ценообразование

Актуальность исследования

В данный момент в теории многоруких бандитов не существует оптимального и точного алгоритма, способного восстанавливать точное распределение цен.

Цель

Предложить онлайн алгоритм многоруких бандитов для нахождения распределений цен на товары-заменители.

Определение

Товары-заменители – товары, находящиеся в одной категории, но разных цен, которые мы и хотим точно определять для увеличения прибыли магазина.

Стохастические многорукие бандиты

В стохастических многоруких бандитах учащийся должен выбрать ручку из конечного множества X в каждый из T раундов. Дергая ручку x в раунде t он получает положительный выигрыш $R_t(x) = r \in \mathcal{R}$ с вероятностью $P(x, r)$. Задача учащегося получить максимальный ожидаемый выигрыш.

Структурированные многорукие бандиты

Распределения выиграшей заранее неизвестны. Известно только то, что $P = (P(r, x))_{r \in \mathcal{R}, x \in X} \in \mathrm{P}$, которое является замкнутым выпуклым множеством с непустой внутренностью. Это знание позволяет ввести некоторую структуру на распределение, а именно
$P(r = 0, x) \geqslant P(r = 0, x') \ \forall x \geqslant x'$.

Литература


Gershkov Alex, Moldovanu Benny, et al.
Revenue maximizing mechanisms with strategic customers and unknown, markovian demand.
In *European Semantic Web Conference*. Springer, 2016.


Weichao Mao, Zhenzhe Zheng, and Fan Wu.
Online pricing for revenue maximization with unknown time discounting valuations.
Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18), pages 440–476.


Parys Van and Golrezaei Negin.
Optimal learning for structured bandits.
arXiv:2011.07738, 2018.

Постановка задачи

Определение

Имеется n товаров заменителей, $p_1, \ldots, p_n$ – распределения товаров, которые мы хотим узнать. Множество покупателей C. Каждый покупатель готов платить за каждый из товаров максимальную цену $\tilde{q}_1(c), \ldots, \tilde{q}_n(c)$. Также введем аппроксимацию стоимостей товаров $q_1(t), \ldots, q_n(t)$ в каждый момент времени t.

Задача

Найти функционал, увеличивая который, сможем получить максимальную прибыль

$$\text{Reg} = \Phi(\mathbf{q}, \tilde{\mathbf{q}}, T, C) \to \min_{\mathbf{q}}$$

Предложение 1

$$\mathbb{E}\,\mathrm{Reg} = T\sum_{c=1}^{|C|} v_1(c) - \mathbb{E}\sum_{t=1}^{T}\sum_{c=1}^{|C|} q_{b_c}(t),\ b_c \in \{1,\dots,n\}.$$

$\mathrm{v}_i(c)$ – пороговая оценка i-го товара для c покупателя.
Также пусть $\forall c \in C \to v_1(c) \geqslant \cdots \geqslant v_n(c)$.
$b_c(t) = \mathrm{argmax}_k\{v_k(c) - q_k(t) :\ q_k(t) \leqslant v_k(c)\}$

Предложение 2

Определение

Политику π назовем равномерно хорошей, если для любого $\alpha > 0$ и для любого распределения $\mathbf{q}$ выполнено

$$\lim \sup_{T \to \infty} \mathbb{E}\, \mathrm{Reg}_\pi(T, \mathbf{q}) / T^\alpha = 0$$

$$\mathrm{Rew}(\mathbf{q}) = \sum_{t=1}^{T} \sum_{c=1}^{|C|} q_{b_c}(t), \;\; \mathrm{Rew}^* = \arg\max_{\mathbf{q}} \mathrm{Rew}(\mathbf{q}).$$

Нижняя граница регрета

Пусть π – произвольная равномерно хорошая политика. Тогда

$$\lim \inf_{T \to \infty} \mathbb{E}\, \mathrm{Reg}_\pi(T, \mathbf{q}) / \log(T) \geqslant C(\mathbf{q})$$

$$C(\mathbf{q}) = \inf_{\mathbf{q}} \Delta(\mathbf{q}), \;\; \Delta(\mathbf{q}) = \mathrm{Rew}^* - \mathrm{Rew}(\mathbf{q})$$

Предложение 3

Главная теорема

Для любого $0 < \varepsilon < 1/n$ существует оптимальная политика π такая, что

$$\lim_{T\to\infty} \sup \mathbb{E}\left[\mathrm{Reg}_\pi(T, \mathbf{q})\right] / \log(T) \leqslant (1+\varepsilon)C(\mathbf{q}) + O(\varepsilon).$$

Алгоритм

DUSA

Предлагается найти двойственную задачу и воспользоваться алгоритмом двойной структуры, который необходимо подстроить под нашу задачу.