Functional multi-armed bandit

И. М. Латыпов

Кафедра Интеллектуальных систем МФТИ

Научный руководитель: к.т.н. Дорн Юрий Владимирович

2025

Проблематика работы

Проблема

Существуют процессы принятия решений с применением экспертов, в которых эксперты улучшаются со временем. Необходимо предложить алгоритм, который будет это учитывать.

Цель

Предложить алгоритм, рассматривающий в качестве экспертов функцию с оптимизатором. Причем для пары функция оптимизатор известна скорость улучшения.

Решение

Алгоритм F-LCB – модификация UCB алгоритма, в ручках которой сидят оптимизаторы.

Пример

Агенты выбора рекламы: есть сайт, на нем показывается реклама.
Также есть несколько агентов для показа рекламы. Каждый из них показывает рекламу и учится на фидбеке от пользователя. Необходимо обеспечить и хороший показ рекламы, и обучение лучших экспертов.

Постановка задачи

Задан набор выпуклых функций $f_i : \mathbb{R}^{n_i} \to \mathbb{R}, i \in \overline{1, K}$. И выпуклые множества $\mathcal{D}_i$. Задано количество раундов T.
Необходимо на каждом раунде выбирать индекс $i_t \in \overline{1, K}$ и вектор $x^{t,i_t} \in \mathcal{D}_{i_t}$.
Алгоритм должен минимизировать регрет:

$$R_O(T) = \sum_{t=1}^{T} \left[f_{i_t}(x^{t,i_t}) - f^* \right]. \tag{1}$$

$f^* = \min_{i \in \overline{1,K}} \min_{x \in D_i} f_i(x)$.

Определение оптимизаторов

Definition

Алгоритм

$$x_{k+1} = \mathcal{A}\left(x_0, \mathcal{O}(x_0), \ldots, x_k, \mathcal{O}(x_k)\right)$$

называется $g(k, \delta)$-ограничивающим, если для любого $k \in \mathbb{N}$ и $\delta > 0$ выполняется неравенство

$$f(x_k) - f(x^*) \leq g(k, \delta)$$

с вероятностью не менее $1 - \delta$.

Если существует функция $g(k)$ такая, что $f(x_k) - f(x^*) \leq g(k)$, то говорят, что алгоритм $\mathcal{A}$ является $g(k)$-ограничивающим.

F-LCB

Algorithm 1 F-LCB algorithm

Require: number of functions K, $g_i(k,\delta)$-bounded optimization method $\mathcal{A}_i$ for $i = 1,\dots,K$, period T, initial estimates $x_0^{\mathcal{P}_1},\dots,x_0^{\mathcal{P}_K}$, parameter δ ($\delta = 0$ for deterministic setup).

1: Run $\mathcal{A}_i$ for each function i ($i = 1,\dots,K$) to compute $x_1^{\mathcal{P}_i} = \mathcal{A}_i(x_0^{\mathcal{P}_i}, \mathcal{O}_{\mathcal{P}_i}(x_0^{\mathcal{P}_i}))$.

2: For each function i ($i = 1,\dots,K$) set $k_i = 1$ and initialize $LCB_i(k_i,\delta) = f_i(x_1^{\mathcal{P}_i}) - g_i(k_i,\delta)$.

3: **for** $t = 1,\dots,T$ **do**

4: Choose function $i_t = \underset{1\le i\le K}{\operatorname{argmin}} LCB_i(k_i,\delta)$.

5: Compute

$$x_{k_{i_t}+1}^{\mathcal{P}_{i_t}} = \mathcal{A}_{i_t}(x_0^{\mathcal{P}_{i_t}}, \mathcal{O}_{\mathcal{P}_{i_t}}(x_0^{\mathcal{P}_{i_t}}),\dots,x_{k_{i_t}}^{\mathcal{P}_{i_t}}, \mathcal{O}_{\mathcal{P}_{i_t}}(x_{k_{i_t}}^{\mathcal{P}_{i_t}})).$$

6: Update LCB index of the played function and preserve others:

$$LCB_{i_t}(k_i+1,\delta) = \begin{cases} LCB_i(k_i,\delta), & i \neq i_t, \\ f_{i_t}(x_{k_{i_t}+1}^{\mathcal{P}_{i_t}}) - g_{i_t}(k_{i_t}+1,\delta), & i = i_t. \end{cases}$$

7: Increase iteration counter for the played arm: $k_{i_t} := k_{i_t} + 1$.

8: **end for**

Рис.: F-LCB алгоритм для выбора лучшей ручки.

Теоретические результаты

Theorem

Пусть $\mathcal{A}_i$ $(i = 1, \ldots, K)$ является $g_i(k)$-ограничивающим алгоритмом. Тогда для алгоритма ***F-LCB*** *для любого $\tau \in \overline{1, T}$ выполнено следующее:*

$$R_O(\tau) \leq \sum_{t=1}^{\tau} g_{i_t}(k_{i_t,t}) = \sum_{i=1}^{K} \sum_{k=1}^{k_{i,\tau}} g_i(k), \tag{2}$$

Здесь $k_{i,t}$ – количество выборов i-ой функции к моменту t.

Теоретические результаты

$$\mathcal{E}_{\text{clean}} \triangleq \{\forall i \in \{1, \ldots, K\}, \forall t \in \{1, \ldots, T\} : f_i(x_{i,t}) - f_i^* \leq g(k_{i,t}, \delta)\}. \tag{3}$$

Assumption

Функции f_i ограничены, т.е. $\max_{1 \leq i \leq K} \max_{x_i \in D_i} f_i(x_i) \leq A$.

Theorem

Пусть $\mathcal{A}_i$ является $g_i(k, \delta)$-ограничивающим алгоритмом для задач $\min_{x \in \mathcal{D}_i} f_i(x)$ *для всех* $1 \leq i \leq K$. *Тогда для регрета $R_O(T)$ алгоритма* ***F-LCB*** *выполнено следующее неравенство:*

$$\mathbb{E}\left[R_O(T)\right] \leq \mathbb{E}\left[\sum_{i=1}^{K} \sum_{t=1}^{k_{i,T}} g_i(t, \delta) \middle| \mathcal{E}_{clean}\right] + \delta K T^2 \cdot A, \tag{4}$$

где $A = \max_{1 \leq i \leq K} \max_{x_i \in D_i} f_i(x_i)$.

Оценки на регрет: детерминированный случай

Таблица: Сводка скоростей сходимости регрета R_O для задачи FMAB в детерминированной постановке. Предполагается, что функции f_i принадлежат одному и тому же классу, а базовые оптимизаторы $\mathcal{A}_i$ одинаковы и указаны во втором столбце. PGD обозначает проекционный градиентный спуск, AGD — ускоренный градиентный спуск, а $\kappa = \frac{L}{\mu}$.

Функция	Базовый оптимизатор	$g(k)$	$R_O(T)$
Выпуклая M-липшицева	PGD	$\frac{RM}{\sqrt{k}}$	$O\left(\sqrt{T \cdot \sum_{i=1}^{K} M_i^2 R_i^2}\right)$
Выпуклая L-гладкая	AGD	$\frac{LR^2}{k^2}$	$O\left(\sum_{i=1}^{K} L_i R_i^2\right)$
μ-сильно выпуклая M-липшицева	PGD	$\frac{M^2}{\mu k}$	$O\left(\left(\sum_{i=1}^{K} \frac{M_i^2}{\mu_i}\right) \log T\right)$
μ-сильно выпуклая L-гладкая	AGD	$R^2 \exp\{-\frac{k}{\sqrt{\kappa}}\}$	$O\left(\sum_{i=1}^{K} \frac{R_i^2}{\exp\left\{\frac{1}{\sqrt{\kappa_i}}\right\}-1}\right)$

Оценки на регрет: стохастика

Assumption

Алгоритм $\mathcal{A}_i$ имеет доступ к несмещённому стохастическому градиентному оракулу, возвращающему $G_i(x, \xi)$. Существуют множество $D_i \subset \mathbb{R}^d$ и значения $\sigma \geq 0$, $\alpha \in (1, 2]$ такие, что для всех $x \in D_i$ выполняется: $\mathbb{E}_\xi \left[\|G_i(x, \xi) - \nabla f_i(x)\|^\alpha \right] \leq \sigma^\alpha$.

Assumption

Для любого x выполняется: $\mathbb{E} \exp \left\{ \frac{\|G_i(x,\xi) - \nabla f_i(x)\|^2}{\sigma_i^2} \right\} \leq 1.$

Таблица: Оценки сожаления для задачи FMAB в стохастическом случае. Алгоритмы SSTM требуют Предположение 2 ($\alpha \in (1, 2]$). AGD требует Предположения 2 ($\alpha = 2$) и 3.

Функция	Базовый оптимизатор	$R_O(T)$
Выпуклая L-гладкая	clipped-SSTM	$O\left(\max\left[KLR^2, \alpha\sigma R K^{1-\frac{1}{\alpha}} T^{\frac{1}{\alpha}} \log(AKT)\right]\right)$
μ-сильно выпуклая, L-гладкая	R-clipped-SSTM	$O\left(\max\left[K\sqrt{\frac{L}{\mu}}, \frac{\sigma^2}{\mu} K^{2\frac{\alpha-1}{\alpha}} T^{\frac{2}{\alpha}-1} \log(AKT)\right]\right)$
μ-сильно выпуклая, M-липшицева	AGD	$O\left(\sqrt{KT}\sigma R \log(AKT)\right)$

Эксперимент

Постановка: В качестве функций рассматриваются нейронки для классификаци на датасете CIFAR10. В качестве оценок на сходимость используются $g_i(k) = \frac{A_i}{\sqrt{k}}$.

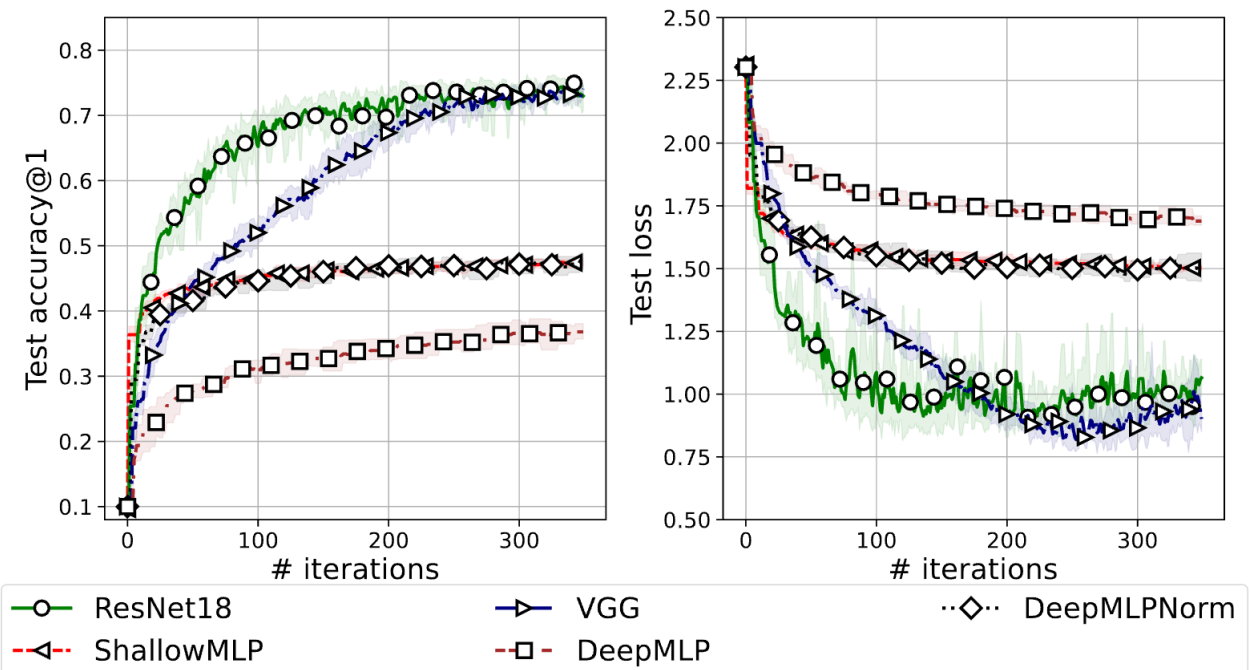

Рис.: Процесс обучения моделей с использованием F-LCB.

Выносится на защиту

1. Поставлена задача функциональных многоруких бандитов.
2. Проанализирован алгоритм F-LCB и получены теоретические результаты.
3. Получены экспериментальные результаты, демонстрирующие работу алгоритма.

Работа подана на NIPS, ожидаются результаты.