РАСПРЕДЕЛЕННОЕ ВЫЧИСЛЕНИЕ БАРИЦЕНТРА ВАСЕРШТЕЙНА

Двинских Д. М.
(*Московский физико-технический институт, Москва;*
Сколковский институт науки и технологий, Москва)

Определим энтропийно-регуляризованное расстояние Васерштейна, порожденное транспортной метрикой на $U(p,q) \triangleq \left\{ X \in \mathbb{R}_{+}^{n\times n} \text{ы} X\mathbf{1} = p, X^T\mathbf{1} = q \right\}$ для двух вероятностных мер $p, q \in S_1(n)$,

(1) $$\mathcal{W}_{\gamma}(p,q) \triangleq \min_{X\in U(p,q)} \left\{ \left\langle M, X \right\rangle + \gamma \sum_{i=1}^{n} \sum_{j=1}^{n} h(X_{ij}) \right\},$$

где M - матрица стоимости перемещения масс с p_i на q_i.

Барицентр Васерштейна для дискретных распределений определяется как решение следующей задачи

(2) $$\min_{p\in S_1(n)} \sum_{i=1}^{m} \lambda_i \mathcal{W}_{\gamma,q_i}(p).$$

Чтобы представить задачу (2) в распределенном виде введем граф $\mathcal{G} = (V, E)$ и определяющую его матрицу Кирхгофа W (матрица коммуникаций)[3]. Введем $p = [p_1^T, \cdots, p_m^T]^T$ и $q = [q_1^T, \cdots, q_m^T]^T$ для каждого $i \in V$, $p_i, q_i \in S_1(n)$ Преобразуя исходную задачу (2), перепишем ее в эквивалентном виде

(3) $$\min_{y} \mathcal{W}_{\gamma,q}^{*}(\sqrt{W}y) := \sum_{i=1}^{m} \mathcal{W}_{\gamma,q_i}^{*}([\sqrt{W}y]_i),$$

где $\mathcal{W}_{\gamma,q_i}^{*}(\sqrt{W}y) = \max_{p_i\in S_1(n)} \left\{ \left\langle \sqrt{W}y, p \right\rangle - \mathcal{W}_{\gamma,q_i}(p) \right\}$ - сопряженная функция. Для решения задачи используем быстрый градиентный метод из [1] в распределенной форме [3].

Алгоритм.

$$\nabla \mathcal{W}^*_{\gamma,q_i}\left([\sqrt{W}y]_i\right)=\sum_{j=1}^{m}\sqrt{W}_{ij}\,p_j^*\left([\sqrt{W}y]_j\right),\tau_k=\frac{2}{k+2},\alpha_{k+1}=\frac{k+2}{2L}$$

Каждому агенту $i\in V$ присваивается распределение q_i, $\tilde{w}_0^i=\tilde{y}_0^i=\tilde{z}_0^i=\mathbf{0}\in\mathbb{R}^n$ и N.

1. For $k=0,1,2,\cdots,N-1$, 2. $\tilde{y}_{k+1}^i=\tau_k\tilde{z}_k^i+(1-\tau_k)\tilde{w}_k^i$,
3. $\tilde{w}_{k+1}^i=\tilde{y}_{k+1}^i-\dfrac{1}{L}\sum_{j=1}^{m}W_{ij}\,p_j^*(\tilde{y}_{k+1}^j)$,
4. $\tilde{z}_{k+1}^i=\tilde{z}_k^i-\alpha_{k+1}\sum_{j=1}^{m}W_{ij}\,p_j^*(\tilde{y}_{k+1}^j)$,
5. Установить $(y_N^*)_i=\tilde{w}_N^i$, $(p_N^*)_i=\sum_{k=0}^{N-1}\dfrac{(k+2)}{N(N+3)}p_i^*(\tilde{y}_{k+1}^i)$.

Результаты. Пусть $\|\nabla\mathcal{W}^*_{\gamma,q}(y)\|_2\le G$ на шаре $B_R(0)$. После

$$N\ge\sqrt{\frac{128G^2m\log n}{\varepsilon^2}\chi(W)}$$

итераций результат алгоритма: $p_N^*=[(p_N^*)_1^T,\cdots,(p_N^*)_m^T]^T$ и $y_N^*=[(y_N^*)_1^T,\cdots,(y_N^*)_m^T]^T$, удовлетворяет следующим свойствам $\mathcal{W}_{0,q}(p_N^*)-\mathcal{W}_{0,q}(p^*)\le\varepsilon$ и $\|\sqrt{W}p_N^*\|_2\le\varepsilon/(2R)$. [2]

Литература

1. ALLEN-ZHU Z., ORECCHIA L. *Linear Coupling: An Ultimate Unification of Gradient and Mirror Descent.* // arXiv preprint arXiv:1407.1537, 2014.
2. URIBE C.A., DVINSKIKH D., DVURECHENSKY P., GASNIKOV A, NEDIC A. *Distributed Computation of Wasserstein Barycenters over Networks* // arXiv preprint arXiv:1803.02933, 2018.
3. URIBE C.A, LEE S., GASNIKOV A., NEDIC. *Optimal Algorithms for Distributed Optimization* // arXiv preprint arXiv:1712.00232, 2017.