Распределенное вычисление барицентров Вассерштейна при меняющихся графах коммуникаций

Юферева О.О. (ИММ УрО РАН)
совместно с Гасниковым А.В., Персияновым М.И. (МФТИ),
Ковалевым Д.А. (King Abdullah University of Science and Technology) и
Двуреченским П.Е. (Weierstrass Institute for Applied Analysis and Stochastics)

Нахождение барицентра Вассерштейна используется при решении практических задач машинного обучения: кластеризации, классификации, распознавании образов и других. В качестве исходных данных могут использоваться картинки, тексты и другие форматы, которые могут быть представлены как вероятностные распределения; их обработка обычно затратна и требует распределенных вычислений. В существующих распределенных методах поиска барицентра Вассерштейна применяется сведение к билинейной седловой задаче, что принципиально использует неизменность лапласиана графа коммуникаций, тогда как в современных задачах коммуникации между устройствами могут быть беспроводными и меняющимися, например, в связи с перемещением устройств в пространстве. В данной работе[1] предложен распределенный метод нахождения барицентра Вассерштейна при меняющихся графах коммуникаций.

В частности, для различных черно-белых картинкок, находящихся на различных устройствах предложенный метод позволяет приблизиться к "средней" (одинаковой для всех устройств) структуре. В эксперименте, представленном на рисунке 1, моделировалось вычисление на 50-ти устройствах соединенных на каждой итерации метода случайным графом Эрдёша–Реньи. Так, на рисунке 1 показано как меняются данные на 7 из 50 устройствах (узлах): в первой строке даны исходные данные датасета черно-белых рукописных цифр "4", в последующих строках показано последовательное изменение данных на устройствах; после 200-ой итерации явно заметна структура "усредненной" цифры "4". На рисунке 2 показан пример генерации меняющихся графов коммуникации.

Более формально, рассматривается задача оптимизации $\frac{1}{m}\sum_{i=1}^{m} f_i(x) \to \min_{x\in S}$, где для $i = 1, \ldots, m$ функции $f_i\colon S \to S$ выпуклы — в данном случае $f_i(x)$ это расстояние Вассерштейна между x и i-ым данным распределением; множество $S \subset \mathbb{R}^d$ выпукло (в данном случае это d-мерный симплекс). Заметим, что не требуется гладкость или строгая выпуклость, и присутствуют ограничения на область определения функций. Кроме того, распределенность вычислений накладывает следующие ограничения. Пусть W_k обозначают графы из m узлов при $k = 1, 2, \ldots$ Предполагается, что m устройств, соединенных на k-ой итерации так же как узлы графа коммуникаций W_k, должны минимизировать сумму $f_i\colon S \to S$. Каждая функция f_i известна только i-му устройству. Каждое устройство на каждой итерации k использует только свои данные и данные своих непосредственных соседей согласно соответствующему графу W_k. Показано, что выход алгоритма приближает условие консенсуса и ε-сходится по функции к оптимальному значению за $\mathcal{O}(\frac{1}{\varepsilon}\ln\frac{1}{\varepsilon})$ итераций.

[1] Yufereva O. et al. Decentralized Computation of Wasserstein Barycenter over Time-Varying Networks //arXiv preprint arXiv:2205.15669. – 2022.

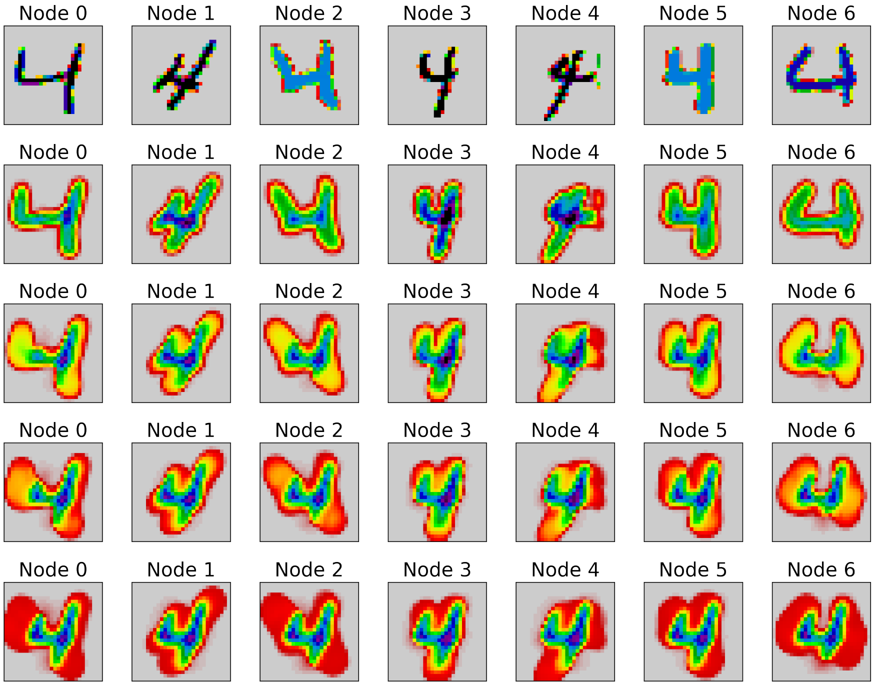

Рис. 1. Изменение локальных данных: исходные данные, преобразованные в вероятностные распределения и результаты распределенных вычислений после 10, 50, 100 и 200 итераций.