

Анализ эффективности децентрализованных алгоритмов оптимизации для задач классификации

Научный руководитель – Манита Лариса Анатольевна

Андреанова А.И.¹, Егорова Э.А.²

1 - Национальный исследовательский университет «Высшая школа экономики», Московский институт электроники и математики им. А.Н. Тихонова, Москва, Россия, *E-mail*: a.andrianova2001@yandex.ru; 2 - Национальный исследовательский университет «Высшая школа экономики», Московский институт электроники и математики им. А.Н. Тихонова, Москва, Россия, *E-mail*: eaegorova_8@edu.hse.ru

Современные методы машинного обучения играют ключевую роль в обработке данных и разработке интеллектуальных систем. Ограничение возможностей передачи данных в связи с трудозатратами и конфиденциальностью является одной из ключевых проблем в современных системах, работающих с распределёнными данными, и требует внедрения новых подходов, таких как федеративное обучение [1-2]. Федеративное обучение (Federated Learning) – это метод, при котором алгоритмы искусственного интеллекта обучаются не на центральном сервере, а на отдельных узлах. В основе федеративных алгоритмов лежат задачи децентрализованной оптимизации, которые представляют собой совокупность какого-либо (в основном, градиентного) метода спуска и консенсусного алгоритма.

В данной работе исследуется эффективность децентрализованных алгоритмов для решения задачи бинарной классификации на основе линейного метода опорных векторов (SVM) с гладкой и негладкой целевой функцией. Рассмотрим N клиентов, у каждого из которых имеется локальный набор данных. Локальные данные для i -ого клиента обозначим как (X_i, Y_i) , где $X_i \in \mathbb{R}^{n_i \times d}$ – матрица признаков для n_i объектов с d -мерным пространством признаков, $Y_i = \{-1, +1\}^{n_i}$ – вектор меток класса. Линейный классификатор определяется набором параметров (w, b) , где w определяет нормаль разделяющей гиперплоскости, b – параметр сдвига. Тогда задача построения единого классификатора сводится к задаче децентрализованной оптимизации следующего вида: $\frac{1}{N} \sum_{i=1}^N Q_i(w, b) \rightarrow \min_{w, b}$, где $Q_i : \mathbb{R}^d \rightarrow \mathbb{R}$ – локальная целевая функция клиента i , $i \in \{1, \dots, N\}$. Отметим, что необходимо найти единый набор параметров (w, b) , который, в общем случае, не является решением задачи минимизации локальной целевой функции. Поэтому возникает задача согласования локальных решений клиентов. Основной метод ее решения – консенсусный алгоритм, в частности, алгоритма ДеГрута.

В данной работе рассмотрены два основных типа обмена информацией: (FDL1) клиенты обмениваются текущими параметрами (w_i, b_i) , (FDL2) клиенты обмениваются градиентами локальных целевых функций (направлениями спуска). В FDL1 алгоритм ДеГрута применяется к наборам локальных параметров, а в FDL2 к текущим направлениям спуска. При этом при передаче информации накладываются случайные факторы $\delta_k^t \in \mathbb{R}^d$, которые представляют собой последовательность независимых одинаково распределённых случайных величин: $\mathbb{E}[\delta_k^t] = 0$, $\mathbb{E}[\|\delta_k^t\|^2] \leq \sigma^2 < \infty$. Результаты моделирования показывают, что FDL2 более устойчив к шуму при решении задачи классификации, чем FDL1. Кроме того, мы исследуем ситуацию, когда объем данных сильно варьируется у разных клиентов (имеется дисбаланс данных). Результаты моделирования демонстрируют, что при дисбалансе данных между клиентами топология, учитывающая объем данных, обеспечивает более низкие значения целевой функции для обоих алгоритмов.

Источники и литература

- 1) Konecny J. Federated optimization: Distributed machine learning for on-device intelligence, arXiv preprint arXiv:1610.02527, 2016, P. 1 - 6.
- 2) Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong, Federated Machine Learning: Concept and Applications //ACM Transactions on Intelligent Systems and Technology (TIST), Volume 10, Issue 2 Article No.: 12, Pages 1 - 19