Оптимизация ансамблевых методов машинного обучения за счёт трансформации целевой переменной

Мельник Фёдор Александрович

Студент

Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия

E-mail: melnik.tedor@gmail.com

Научный руководитель *— д.ф-м.н. проф. Сенько Олег Валентинович*

Решающие леса [2] наравне с градиентным бустингом [3] являются классическим методом машинного обучения, решать задачи регрессии и классификации. Однако у этого метода есть недостаток - при высокой корреляции между различными деревьями общее качество работы алгоритма будет невысоким. Одним из возможных способов решения данной проблемы является замена целевой переменной на величину, зависящую как от изначальной целевой переменной, так и от прогнозов ранее обученных деревьев.

В работе [1] был предложен следующий метод: вместо оптимизации функционала $L(y, A(x)) = \frac{1}{m}\sum\limits_{j=1}^{m}(y_j - \overline{A}_h(x))^2$, как в обычных решающих деревьях, можно оптимизировать величину $D(y, A(x)) := \frac{1}{km}\sum\limits_{j=1}^{m}(y_j - A_k(x_j))^2 - \frac{\mu}{km}\sum\limits_{j=1}^{m}(\overline{A}_k(x_j) - A_k(x_j))^2$, что приводит к тому, что на каждом шаге предсказывается величина $t_j = \frac{1}{1-\Theta}y_j - \frac{\Theta}{1-\Theta}\overline{A}_{k-1}(x_j)$, где $\Theta = \mu\frac{(k-1)^2}{k^2}$.

На основе этого данного метода можно предложить аналогичные подходы. Так, если минимизировать функции $LogLoss(y, A_k) - \mu LogLoss(A_k, \overline{A}_{k-1})$ (для задачи классификации) или $MSE(y, A_k) - \mu MSE(A_k, \overline{A}_{k-1})$ (для задачи регрессии). В этом случае целевая переменная принимает вид $\frac{y-\mu\overline{A}_{k-1}}{1-\mu}$. Данные метод показал свою высокую эффективность в задачах классификации и при решении задачи ранжирования путём сведения к задаче классификации.

Однако такой метод часто может давать крайне низкие результаты в тех случаях, когда μ близко к 1. Для решения этой задачи были испробованы следующие методы:

- замена отдельного дерева на градиентный бустинг из небольшого числа деревьев;

- прогнозирование $\frac{y-\mu\overline{A}_{k-1}+\beta\mathbb{E}y}{1-\mu+\beta}$, где β - произвольный параметр;
- выбор веса для каждого дерева в зависимости от его прогнозов;
- замена целевой переменной в нескольких первых деревьях на y;
- использование для построения дерева признаки, выбирающиеся из общего множества с весами, зависящими от их важности (а не абсолютно случайно, как в случае классического решающего леса).

Каждый из подходов позволял дополнительно повысить качество задач регрессии и классификации на отдельных наборах данных, что в совокупности позволяет добиться в среднем более высокого качества, а также повысить стабильность прогнозов и снизить влияние точного подбора гиперпараметров на итоговый результат

Литература

1. Sen'ko, O., Dokukin, A., Kiselyova, N., Dudarev, V. and Kuznetsova, Yu. (2023) 'New two-level ensemble method and its application to chemical compounds properties prediction', *Lobachevskii Journal of Mathematics*, 44, pp. 188–197. doi:10.1134/S1995080223010341.
2. Breiman, L. (2001) 'Random forests', Machine Learning, 45(1), pp. 5–32. doi:10.1023/A:1010933404324
3. Friedman, J.H. (2001) 'Greedy function approximation: A gradient boosting machine', Annals of Statistics, 29(5), pp. 1189–1232.