

МЕТОДЫ ПОСТРОЕНИЯ АНСАМБЛЕЙ НАД НАИБОЛЕЕ РАЗНООБРАЗНЫМИ АЛГОРИТМАМИ С ИСПОЛЬЗОВАНИЕМ СМЕЩЕННОЙ ФУНКЦИИ ПОТЕРЬ

Саенко Иван Александрович

Студент

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

E-mail: saenkoia@ph.msu.ru

Научный руководитель — Сенько Олег Валентинович

Среди большого количества методов машинного обучения важную роль играют те, в которых для построения итогового прогноза используется множество простых и не очень точных алгоритмов (слабых предикторов). Например, к подобным методам могут быть отнесены алгоритмы градиентного бустинга и случайного леса.

Самое главное в работе таких методов - способ построения простых алгоритмов. Хороший способ увеличения обобщающей способности итогового прогноза - строить эти простые алгоритмы таким образом, чтобы они имели максимально отличающиеся прогнозы. Например, в случайном лесе это достигается благодаря разнообразию выборок, на которых строятся деревья (там они выбираются с помощью бутстрэпа).

В данной работе рассматриваются несколько новых методов построения многослойных ансамблей на основе решающих деревьев, построенных специальным образом на основе предсказаний предыдущих деревьев.

Первый из предлагаемых методов - метод сдвигаемой целевой переменной. Основная идея метода состоит в том, что обучение каждого следующего простого алгоритма происходит на измененной целевой переменной: конкретнее, если реальная целевая переменная - это y и уже имеются m построенных алгоритмов, каждый из которых выдал предсказание \hat{y}_i , то следующий $(m + 1)$ -ый алгоритм должен предсказывать $y - \alpha * \bar{y}$, где $\bar{y} = \sum_{i=1}^m \hat{y}_i$. Полученные алгоритмы можно как усреднить, так и использовать в качестве базовых алгоритмов для второго уровня ансамблирования, который потенциально может привести к улучшению качества работы алгоритма. В качестве ансамблирующих алгоритмов были рассмотрены случайный лес, эластичная сеть и набор эластичных сетей.

Второй рассматриваемый метод - метод корректирующего дерева. В данном методе оптимизируется следующий функционал: $Q = (\hat{Y} - Y)^2 - \alpha * (\hat{Y} - \hat{Y}_{ens})^2$, где Y - это целевая переменная, \hat{Y} -

предсказание текущего алгоритма, \hat{Y}_{ens} - предсказание ансамбля на предыдущем шаге, α - гиперпараметр. Важно отметить, что функционал Q на каждом шаге оптимизируется именно по текущему алгоритму. Таким образом происходит попытка получать простые алгоритмы, которые одновременно хорошо и предсказывают целевую переменную сами по себе, и находятся как можно дальше от имеющегося ансамбля. Для построения очередного базового алгоритма нужно сделать два шага:

1) На первом шаге с помощью метода бутстрэпа из исходной обучающей выборки X генерируется выборка, на которой обучается базовый алгоритм A_0 ; предсказания этого алгоритма запишем как \hat{Y}_0 .

2) На втором шаге сначала строим алгоритм (например, решающее дерево) A_1 , которое обучается на всей выборке X на целевую переменную $\frac{\partial Q}{\partial(\hat{Y} - \hat{Y}_0)}$, где Q определена ранее. Имея алгоритмы A_0 и A_1 , добавляем в ансамбль $A_{new} = A_0 + A_1$.

Для этого метода также рассмотрены несколько способов построения второго слоя. В экспериментах проверялись ансамблирование усреднением, случайным лесом и градиентным бустингом.

Результаты экспериментов показывают, что в большинстве случаев предложенные алгоритмы показывают более высокое качество, чем стандартные алгоритмы градиентного бустинга и случайного леса.

Литература

1. Журавлев Ю. И., Сенько О. В., Докукин А. А., Киселева Н. Н., Саенко И. А. Двухуровневый метод регрессионного анализа, использующий ансамбли деревьев с оптимальной дивергенцией // Доклады Российской академии наук. Математика, информатика, процессы управления. 2021. Т. 499, № 1. С. 63–66.
2. Breiman L. Bagging predictors // Machine learning. 1996. Vol. 24, № 2. P. 123–140.
3. Zou H., Hastie T. Regularization and variable selection via the elastic net // Journal of the Royal Statistical Society, Series B. 2005. P. 301–320.