

# Эффективное вычисление SHAP значений для кусочно-линейных решающих деревьев

А.К. Гурьянов<sup>1</sup>

<sup>1</sup>Московский государственный университет им. М.В. Ломоносова, Ленинские горы, 1, Москва, Россия, 119991

## Аннотация

Интерпретируемость моделей машинного обучения важна во многих прикладных задачах анализа данных. В последние годы стала популярной универсальная парадигма интерпретации SHAP на основе значений Шепли. Однако для интерпретации с помощью SHAP значений напрямую необходимо рассчитывать предсказание модели экспоненциальное от размера признакового пространства количество раз, что приводит к запретительной сложности такого метода расчета для сложных моделей, например для нейронных сетей или ансамблей градиентного бустинга. Было показано, что существуют эффективные алгоритмы вычисления SHAP значений для решающих деревьев и аддитивных ансамблей на их основе, использующие структуру деревьев для теоретически обоснованного расчета. Один из новых интересных классов моделей машинного обучения — это кусочно-линейные решающие деревья и ансамбли градиентного бустинга над ними. В данной работе был предложен эффективный алгоритм вычислений SHAP значений для кусочно-линейных решающих деревьев и аддитивных ансамблей на их основе.

## Ключевые слова

Машинное обучение, Интерпретация моделей машинного обучения, SHAP, Кусочно-линейные решающие деревья, Градиентный бустинг

## 1. Введение

В последние годы алгоритмы машинного обучения используются во все большем кругу отраслей. Актуальной задачей является задача их интерпретации. Одна из наиболее интересных технологий для интерпретации, разработанная в последние годы — это SHAP [1] значения моделей машинного обучения, которые позволяют оценить вклад каждого фактора в итоговое решение, и при этом обладают рядом важных и полезных свойств. Для произвольной модели машинного обучения вычисление SHAP значений является вычислительно затратным, но для ряда моделей машинного обучения, в том числе линейных моделей и ансамблей градиентного бустинга на основе решающих деревьев [2][3], были предложены способы эффективного расчета [4].

Интересной новой моделью машинного обучения являются кусочно-линейные решающие деревья [5][6][7]. Подобные модели, являющиеся усложнением стандартных решающих деревьев, позволяют добиваться лучшего качества предсказания без значительного увеличения времени предсказания, что является важным фактором в некоторых прикладных областях. В последние годы были предложены ряд эффективных способов обучения подобных моделей, в том числе в парадигме градиентного бустинга. Алгоритмы для эффективного вычисления SHAP значений, разработанные для стандартных решающих деревьев, не подходят для расчета SHAP значений подобных новых моделей, из-за чего интерпретация кусочно-линейных решающих деревьев и бустинга над ними является вычислительно дорогой.

В рамках данной работы предоставлено обоснование корректности эффективного алгоритма для расчета SHAP значений стандартных решающих деревьев, и предложен новый алгоритм для эффективного расчета SHAP значений кусочно-линейных деревьев, позволяющий сократить алгоритмическую сложность интерпретации.

## 2. SHAP-значения

Один из новых способов интерпретации моделей машинного обучения — это т.н SHAP значения. Для вычисления SHAP значений для произвольной модели  $f$  в объекте  $x$  вводится функция  $f_x(S)$ , где  $S$  - это некоторое подмножество признаков пространства объекта  $x$ , которая определяет математическое ожидание предсказания модели в объекте  $x$  при условии, что модели известны значения признаков только из множества  $S$ . Вклад  $\phi_i$  признака  $i$  в предсказание модели определяется как математическое ожидание прироста предсказания модели от добавления признака  $i$  в множество известных модели признаков, взвешенное по всем возможным подмножествам признаков пространства. Для вычисления SHAP значений модели для одного объекта выборки напрямую необходимо рассчитать условное математическое ожидание предсказания модели по всем доступным множествам признаков. Наивный алгоритм для решающих деревьев на основе этой логики имеет алгоритмическую сложность  $O(L2^M)$ , где  $L$  — количество листов в дереве,  $m$  — количество признаков.

## 3. Кусочно-линейные решающие деревья

Один из интересных алгоритмов машинного обучения — это кусочно-линейные решающие деревья. В таких деревьях решений предсказаниями в листьях являются линейные функции над признаковым пространством. За последние годы вышел ряд работ, предлагающих эффективные способы обучения таких моделей, и показывающие конкурентноспособное качество по сравнению с популярными решениями на основе стандартных решающих деревьев.

В данной работе предлагается выделить три класса кусочно-линейных решающих деревьев:

- Общие кусочно-линейные решающие деревья;
- Кусочно-линейные решающие деревья, в которых предсказания в листьях являются линейными функциями только от признаков, участвующих в решающих правилах на пути к этим листьям;
- Кусочно-линейные решающие деревья второго класса с дополнительными ограничениями на весовые коэффициенты у линейных функций в листьях, предложены в статье [5].

Для каждого из выделенных классов предложены различные алгоритмы, позволяющие эффективно рассчитать SHAP-значения за полиномиальное время от количества признаков, количества листьев в дереве и глубины дерева. Расчет SHAP значений для ансамблей градиентного бустинга кусочно-линейных решающих деревьев заключается в независимом расчете SHAP значений для каждого из деревьев ансамбля и суммировании полученных значений для каждого из признаков. Подобное обобщение алгоритма расчета для ансамблей верно благодаря свойству аддитивности SHAP значений.

Для одного из классов кусочно-линейных решающих деревьев предложенный алгоритм реализован в виде программного модуля, корректность алгоритма и расчета SHAP значений продемонстрирована на открытом наборе данных HIGGS [8].

## 4. Заключение

Данная работа приводит теоретическое и практическое обоснование использованию SHAP-значений в моделях градиентного бустинга над кусочно-линейными решающими деревьями.

## 5. Литература

- [1] Lundberg, S.M. A unified approach to interpreting model predictions / S.M. Lundberg, S.-I. Lee // Advances in neural information processing systems. – 2017.
- [2] Quinlan, J.R. Induction of decision trees // Machine learning. – 1986. – Vol. 1.1. – P. 81-106.
- [3] Chen, T. Xgboost: A scalable tree boosting system / T. Chen, C. Guestrin // Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. – 2016.

- [4] Lundberg, S.M. Consistent individualized feature attribution for tree ensembles / S.M. Lundberg, G.G. Erion, S.-I. Lee // ArXiv preprint: 1802.03888. – 2018.
- [5] Guryanov, A. Histogram-Based Algorithm for Building Gradient Boosting Ensembles of Piecewise Linear Decision Trees // International Conference on Analysis of Images, Social Networks and Texts. – 2019.
- [6] de Vito, L. LinXGBoost: extension of XGBoost to generalized local linear models // ArXiv preprint: 1710.03634. – 2017.
- [7] Shi, Y. Gradient boosting with piece-wise linear regression trees / Y. Shi, J. Li, Zh. Li // ArXiv preprint: 1802.05640. – 2018.
- [8] Baldi, P. Searching for Exotic Particles in High-energy Physics with Deep Learning / P. Baldi, P. Sadowski, D. Whiteson // Nature Communications. – 2014. – Vol. 5.