

**КРИТИЧЕСКИЙ ОБЗОР МЕТОДОВ АНАЛИЗА
ВЫЖИВАЕМОСТИ НА ОСНОВЕ БУСТИНГ
АНСАМБЛЕЙ**

Васильев Юлий Алексеевич

Аспирант

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

E-mail: iuliivasilev@gmail.com

Научный руководитель — Петровский Михаил Игоревич

Задачей анализа выживаемости является прогнозирование времени и вероятности до наступления определенного события. Анализ событий активно применяется в инженерных науках, здравоохранении, кредитном скоринге и т.д.

В общем виде задача анализа выживаемости сводится к построению двух функций [1]. Функция выживания (survival function) определяет вероятность не наступления события по истечении определенного времени. Функция риска (hazard function) определяет вероятность события в конкретный момент времени, при условии, что событие не наступило ранее.

Основной проблемой анализа выживаемости является наличие цензурированных наблюдений, для которых событие не наблюдалось в течение всего исследования (или они были потеряны).

Также, задача анализа выживаемости может быть представлена в непрерывном и дискретном виде [2]. В задаче непрерывного времени рассматривается вся временная шкала, но задаются строгие допущения о распределении целевых переменных. Для задачи дискретного времени временная шкала дискретизируется по заданным временным интервалам. Такой подход требует меньше допущений, но приводит к потере информации, сильно влияет на точность и вычислительную сложность построения модели. Наконец, задача дискретного времени накладывает серьезное функциональное ограничение на допустимые моменты времени для прогнозирования.

На данный момент, наиболее популярными методами машинного обучения являются ансамбли деревьев решений. В частности, градиентный бустинг [3] основан на построении композиции алгоритмов, так что каждый следующий алгоритм стремится компенсировать недостатки всех предыдущих алгоритмов.

Основные характеристики градиентного бустинга: простота, универсальность, гибкость к модификациям и высокая обобщающая

способность. Однако, градиентный бустинг позволяет работать только с одной целевой переменной, в то время как в анализе выживаемости существуют 2 различные по логике переменные: время события и индикатор цензурирования.

По результатам проведенного обзора литературы и открытых реализаций, могут быть выделены 3 основных недостатка существующих бустинг ансамблей анализа выживаемости. Во-первых, ограничение функциональности приводит к невозможности прогнозирования функций выживания и риска [4]. Во-вторых, сводя задачу к классификации и регрессии с последующим построением функций, вводится дополнительное допущение о пропорциональности рисков, которое не применимо на практике [5]. Наконец, сводя задачу к построению множеству дискретных моделей, повышается вычислительная сложность построения и прогнозирования модели [6].

Для решения данных недостатков возможно построение векторного градиентного бустинга для дискретной задачи анализа выживаемости, целью которого является прогнозирование функции риска. Альтернативным подходом является использование адаптивных ансамблей деревьев выживаемости, основанных на последовательном изменении весов значимости наблюдений для построения последующих алгоритмов.

Литература

1. Wang P., Li Y., Reddy C. K. Machine learning for survival analysis: A survey //ACM Computing Surveys (CSUR). – 2019. – Т. 51. – №. 6. – С. 1-36.
2. Kvanme H., Borgan Ø. Continuous and discrete-time survival prediction with neural networks //arXiv preprint arXiv:1910.06724. – 2019.
3. Friedman J. H. Stochastic gradient boosting //Computational statistics & data analysis. – 2002. – Т. 38. – №. 4. – С. 367-378.
4. Chen Y. et al. A gradient boosting algorithm for survival analysis via direct optimization of concordance index //Computational and mathematical methods in medicine. – 2013. – Т. 2013.
5. Nguyen N. P. Gradient Boosting for Survival Analysis with Applications in Oncology. – University of South Florida, 2019.
6. Bai M., Zheng Y., Shen Y. Gradient boosting survival tree with applications in credit scoring //Journal of the Operational Research Society. – 2022. – Т. 73. – №. 1. – С. 39-55.