

УДК 314.48

DOI: [10.26102/2310-6018/2021.35.4.028](https://doi.org/10.26102/2310-6018/2021.35.4.028)

Применение метода «временного квилтинга» для анализа выживаемости после инфаркта миокарда

М.А. Фирюлина

*Воронежский государственный университет,
Воронеж, Российская Федерация*

Резюме: Значимость анализа выживаемости в задачах медицинского характера привела к развитию множества подходов к моделированию функции выживаемости. Модели, построенные с помощью различных методов машинного обучения, имеют сильные и слабые стороны с точки зрения различительной производительности и возможностей калибровки, но ни одна модель не является лучшей для всех наборов данных или даже для всех временных горизонтов в пределах одного набора данных. Актуальность исследования обусловлена тем, что не всегда базовые модели и ансамблевые подходы позволяют построить хорошую модель выживаемости для разных временных горизонтов. В связи с этим, данная статья направлена на описание применения нового подхода, который объединяет различные базовые модели для создания достоверной функции выживаемости, которая предоставляет возможности для настройки и имеет хорошие дискриминантные характеристики в различных временных горизонтах. В ходе исследования было рассмотрено шесть базовых моделей анализа выживаемости после инфаркта миокарда: непараметрические методы (модель пропорциональных рисков Кокса, модель пропорциональных рисков Кокса с использованием гребневой регрессии), параметрические модели (логистическая модель нормального распределения, логистическая модель экспоненциального распределения, метода распределения Вейбулла) и ансамблевая модель (случайный лес). Ведущим подходом к решению данной проблемы является применение усовершенствованного метода – временного квилтинга. В статье показано сравнение данного подхода с базовыми относительно точности и оценки калибровки модели. По результатам исследования выявлено, что наиболее эффективной оказалась модель временного квилтинга, а наименее эффективной – модель случайного леса. Поскольку усовершенствованный подход автоматически находит аппроксимацию наилучшей модели выживания, он дает возможность клиницистам избавиться от траты времени на поиск одной конкретной модели выживания для каждого набора данных и для каждого интересующего временного горизонта.

Ключевые слова: машинное обучение, анализ выживаемости, временной квилтинг, байесовская оптимизация, инфаркт миокарда.

Благодарности: Исследование выполнено при поддержке РФФИ, проект 20-37-90029 Аспиранты.

Для цитирования: Фирюлина М.А. Применение метода «временного квилтинга» для анализа выживаемости после инфаркта миокарда. *Моделирование, оптимизация и информационные технологии*. 2021;9(4). Доступно по: <https://moitvvt.ru/ru/journal/pdf?id=1080> DOI: 10.26102/2310-6018/2021.35.4.028

Application of “temporal quilting” method for survival analysis after myocardial infarction

M.A. Firyulina

*Voronezh State University,
Voronezh, Russian Federation*

Abstract: The importance of survival analysis in medical problems has led to development of a variety of approaches to modeling the survival function. Models built with various machine learning methods have strengths and weaknesses in terms of differential performance and calibration capabilities, but no model is most suitable for all datasets or even all-time horizons within a single dataset. The relevance of the research is due to the fact that basic models and ensemble approaches do not always make it possible to build a proper survival model for different time horizons. Because of that, this article aims to outline the application of a new approach that combines various basic models to create a reliable survival function, providing opportunities for fine-tuning and having good discriminant characteristics in different time horizons. During the course of the study, six basic models for analyzing survival after myocardial infarction were described: nonparametric methods (Cox proportional hazards model, Cox proportional hazards model using ridge regression), parametric models (logistic normal distribution model, logistic exponential distribution model, Weibull distribution method) and ensemble model (random forest). The principal approach to solving this problem is the use of an improved method – temporal quilting. In this study, the aforementioned approach is compared to basic methods in relation to accuracy and assessment of model calibration. The research results have revealed that ‘temporal quilting’ model is the most efficient while random forest model appears to be the least efficient. Since the enhanced approach automatically finds the approximation of the best-suited survival model, it enables clinicians to reduce time spent on the search for one specific survival model for each dataset as well as for each relevant all-time horizon.

Keywords: machine learning, survival analysis, temporal Quilting, Bayesian optimization, myocardial infarction.

Acknowledgments: The study was carried out with the financial support of the Russian Foundation for Basic Research within the framework of the scientific project No. 20-37-90029 Postgraduates.

For citation: Firyulina M.A. Application of "temporal quilting" method for survival analysis after myocardial infarction. *Modeling, Optimization and Information Technology*. 2021;9(4). Available from: <https://moitvvt.ru/ru/journal/pdf?id=1080> DOI: 10.26102/2310-6018/2021.35.4.028 (In Russ).

Введение

Анализ выживаемости (оценка времени до наступления определенного события) играет важную роль во многих сферах, особенно в медицине. Важность анализа выживаемости побудила к разработке различных подходов к моделированию функции выживаемости (представляющую вероятность выживания в течение заданного времени как функцию нескольких независимых переменных). Параметрический и полупараметрический подходы создают модели, основанные на конкретных предположениях об истинном базовом распределении; непараметрические подходы предлагают построение модели локально вокруг имеющихся данных. Модели, созданные с помощью этих различных подходов, имеют разные сильные и слабые стороны как с точки зрения различительной производительности, так и с точки зрения возможностей калибровки, а их относительная производительность варьируется на разных наборах данных и на разных временных горизонтах в пределах одного набора данных. В частности, ни одна модель не является лучшей для всех наборов данных.

К числу параметрических относятся модели выживаемости, основанные на модели пропорциональных рисков Кокса [1], множество расширений этой модели [2, 3, 4] и модель ускоренного времени отказа (AFT), основанная на использовании распределения Вейбулла [5].

К непараметрическим относятся модели, основанные на ансамблях деревьев выживания, построенные с помощью бэггинга или бустинга [6, 7], методы глубокого обучения [8]. В целом непараметрические модели обеспечивают лучшие прогнозы выживаемости, чем параметрические модели, когда истинное базовое распределение неизвестно или указано неверно. Однако непараметрические модели часто дают

неточные прогнозы на временных горизонтах, для которых количество субъектов в наборе данных, которые находятся «в группе риска», невелико [9].

Эффективность модели выживания следует оценивать по тому, насколько хорошо модель прогнозирует риски наступления события и насколько хорошо модель откалибрована (согласуется с экспериментальными данными).

Адекватная оценка риска смертности способствует своевременному принятию терапевтических методов для улучшения состояния пациента. Несмотря на то, что оценке прогноза жизни больных, перенесших ИМ, и влияющих на него факторов посвящено уже большое количество исследований, итоговые результаты могут быть улучшены с помощью некоторых методов. В задачах медицины важно выяснить, какая из моделей точнее оценивает вероятности (как визуально, так и численно) и как «подправить» имеющуюся модель, чтобы увеличить точность.

На предыдущем этапе предлагаемого в данной статье исследования для анализируемого набора данных уже был проведен анализ выживаемости [10]. Но прогноз производился по отдельности с помощью базовых моделей машинного обучения и классических моделей выживаемости Каплана-Майера и Кокса. В данной статье рассматривается шесть базовых моделей для анализа выживаемости, а также новый подход к решению этих проблем. Этот подход представляет собой объединение различных базовых моделей выживания, чтобы создать функцию, которая является как хорошо дискриминирующей (разделяющей пациентов с высоким и низким риском смертности после инфаркта миокарда), так и хорошо откалиброванной. Поскольку объединяются базовые модели в соответствии с весами, которые меняются во времени, модель называется «временной квилтинг», а сам метод получения модели – «одеяло выживаемости». [11].

Большинство популярных методов машинного обучения, реализованных в различных библиотеках, дает возможность оценить вероятность попадания в тот или иной класс (например, LogisticRegression, SVC, RandomForest, XGBoost, LightGBM, CatBoost). Но многие исследования показали, что самые популярные прогностические модели недостаточно хорошо откалиброваны [12, 13]. Целью данного исследования является сравнение нескольких методов анализа выживаемости для оценки калибровки модели и выбора наиболее эффективной.

Материалы и методы

Описание исходных данных

Для анализа использовалась полная выборка пациентов, поступивших в 2015-2017 годах в больницы Воронежской области с диагнозом ИМ [10]. Всего в исследовании было рассмотрено 11457 случаев инфаркта миокарда, из них 2025 (17.7 %) случаев с летальным исходом и 9432 (82.3 %) – выжившие пациенты. Анализ проводился по следующим факторам: пол, возрастная группа, артериальная гипертензия (АГ), является ли инфаркт миокарда повторным (ИМ), сахарный диабет (СД), фибрилляция предсердий (ФП), острое нарушение мозгового кровообращения (ОНМК), хроническая обструктивная болезнь легких (ХОБЛ), хроническая сердечно-сосудистая недостаточность (ХСН), локализация, тяжесть по KILLIP и проводилась ли пациенту тромболитическая терапия (ТЛТ) и чрескожные коронарные вмешательства (ЧКВ). Единственная числовая переменная в исходной выборке – возраст. Средний возраст пациентов 66 ± 12 лет. Для упрощения интерпретации полученных результатов возраст пациентов был разбит на 8 категорий.

Используемые методы машинного обучения

Анализ выживаемости – это особый тип задачи регрессии. Он отличается от традиционной регрессии тем, что часть обучающих данных цензурирована, то есть их можно наблюдать только частично. Для пациентов, у которых неизвестен окончательный результат ввиду выбывания из исследования, нельзя указать, наступило ли для них впоследствии интересующее нас событие (например, смерть). Единственная достоверная информация, доступная для подобных пациентов, заключается в том, что у них не было этого события до последнего наблюдения. Следовательно, анализ выживаемости требует моделей, которые учитывают эту уникальную характеристику такого набора данных.

Модель пропорциональных рисков Кокса (CoxPH) позволяет оценить выживаемость и функцию совокупных рисков при наличии дополнительных ковариат (независимых признаков). Это возможно, поскольку предполагается, что существует базовая функция риска и что ковариаты изменяют риск только пропорционально. Другими словами, предполагается, что соотношение риска наступления события у любых двух различных пациентов остается постоянным с течением времени.

CoxPHRidge – это модификация стандартной модели CoxPH, но с использованием гребневой регрессии (подхода, предохраняющего модель от переобучения, то есть чрезмерной подгонки под данные обучающей выборки, снижающей способность модели к обобщению). Для построения этих моделей были выбраны линейные модели из библиотеки `sksurv` (`sksurv.linear_model.CoxPHSurvivalAnalysis`). `scikit-Survival` – это модуль Python для анализа выживаемости, построенный на основе `scikit-learn`. Он позволяет проводить анализ выживаемости, используя возможности `scikit-learn`, например, для предварительной обработки или перекрестной проверки.

`Lifelines` – это еще одна полная библиотека анализа выживаемости на языке Python. Основные ее преимущества – это возможность обработки интервальных цензурированных данных. Также эта библиотека содержит наиболее популярные параметрические, полупараметрические и непараметрические модели. Из данной библиотеки использовались модели выживаемости с использованием распределения Вейбулла (Weibull), логарифмически нормального (LogNormal) и экспоненциального (Exponential) распределения (`WeibullAFTFitter`, `LogNormalAFTFitter`, `LogLogisticAFTFitter`). Данные модели относятся к типу параметрических. Все три модели относятся к классу AFT. Модель ускоренного времени отказа – это параметрическая модель, которая предоставляет альтернативу модели Кокса. В то время как модель пропорциональных рисков предполагает, что эффект ковариаты заключается в умножении риска на некоторую константу, модель AFT предполагает, что эффект ковариаты заключается в ускорении или замедлении срока каким-то постоянным значением. Это особенно привлекательно в техническом контексте, где «болезнь» является результатом некоторого механического процесса с известной последовательностью промежуточных стадий. Это можно интерпретировать как замедление или ускорение выполнения функции выживания. Между собой модели различаются способом моделирования времени до наступления события. Отношение рисков не является частью этой модели AFT, поэтому здесь нет подобной интерпретации.

`RandomSurvForest` – это метаоценка, которая подбирает несколько деревьев выживания для различных подвыборок набора данных и использует усреднение для повышения точности прогнозов и контроля чрезмерной подгонки.

Четко сформировать преимущества и недостатки каждого метода в сравнении с другим сложно. По производительности они не уступают друг другу, а точность зависит от контекста поставленной задачи и исходного набора данных.

Обобщающий метод, который используется в данном исследовании, это метод временного квилтинга, который построен на модели одеяла выживания [11]. Принцип данного метода состоит в объединении различных моделей выживания в зависимости от времени, чтобы обеспечить как дискриминирующую, так и прогностическую точность. В данном подходе используются веса, зависящие от времени, для построения достоверной функции выживания, которая хорошо откалибрована и обеспечивает хорошие дискриминирующие оценки риска в различных временных горизонтах. Ядром метода является алгоритм последовательного конфигурирования весов по сетке временных интервалов. Чтобы сделать проблему решаемой, применяется байесовская оптимизация, которая моделирует показатели точности дискриминации и калибровки как выходы черного ящика, входные данные которого представляют собой массив весов (по разным временным горизонтам), а выходными данными являются достигнутая соответствующая точность. Байесовская оптимизация – это метод глобальной оптимизации для неизвестной функции (черного ящика) с шумом. Этот метод часто используется для решения задачи оптимизации гиперпараметров в задачах машинного обучения. Байесовская оптимизация строит стохастическую модель, входами которой являются значения гиперпараметров, а выходами – показатели точности такой модели на тестовом множестве. На практике байесовская оптимизация показывает лучшие результаты по сравнению с другим часто используемым методом – простым перебором значений гиперпараметров по сетке, ввиду возможности быстрее предсказывать результат изменения гиперпараметров.

На основе найденного массива весов метод временного квилтинга создает единую прогностическую модель – одеяло выживания. Также данный подход дает возможность избавиться от выбора одной конкретной модели выживания для каждого набора данных и для каждого интересующего временного горизонта. На Рисунке 1 проиллюстрирован принцип временного квилтинга.

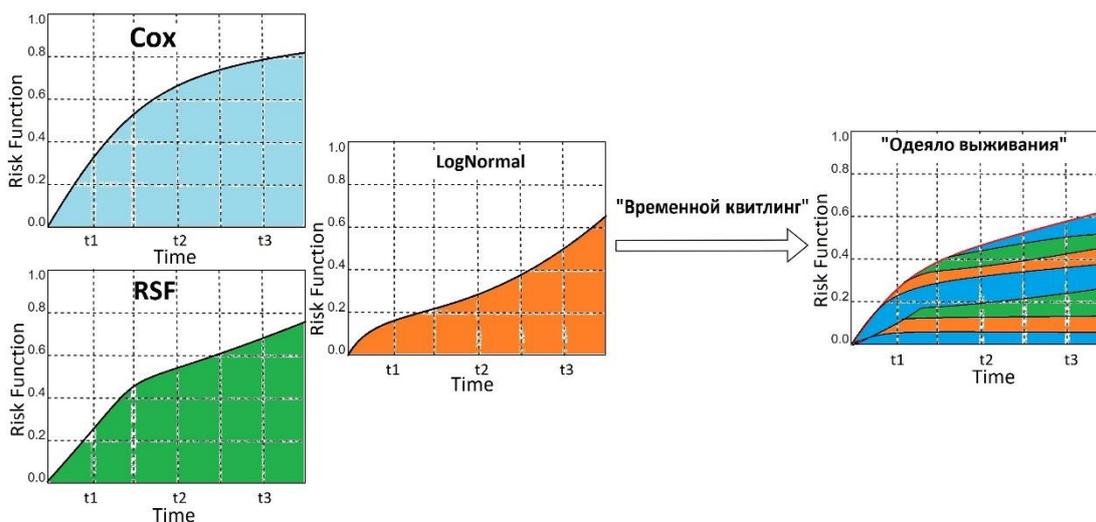


Рисунок 1 – Принцип работы модели временного квилтинга
 Figure 1 – Model of temporal quilting’s work

Таким образом, новый подход к анализу выживания создает изменяющиеся во времени ансамбли существующих моделей выживания. Одеяла для выживания

используют существующие модели, придавая им больший вес во временных интервалах, где эти модели обеспечивают лучшую производительность, и меньший вес во временных интервалах, когда эти модели обеспечивают менее хорошую производительность. Одним из достоинств подхода является то, что имеется возможность адаптироваться к использованию других моделей выживания.

Введем некоторые обозначения. Пусть для каждого пациента i определен вектор входных предикторов $x_i \in X$ (X – множество всех предикторов), время до наступления события $T_i \in \mathbb{R}$, $C_i \in \mathbb{R}$ время до момента нахождения под наблюдением (цензурированные данные) и показатель $\Delta = 1$ если данные полные и $\Delta = 0$, если данные цензурированные (то есть пациент выбыл из исследования и окончательный результат неизвестен).

Таким образом, получаем множество данных D для N пациентов, $D = \{(x_i, \Delta_i T_i + (1 - \Delta_i)C_i, \Delta_i)\}_{i=1}^N$. Цель данного метода в оценке функции риска для пациента $R \rightarrow [0,1]$

$$R(t|x) = \mathbb{P}(T \leq t|x), \quad (1)$$

которая является вероятностью того, что событие произойдет до определенного момента времени с учетом входных характеристик пациента. Если необходимо рассматривать обратное явление, что событие не наступит, то достаточно рассмотреть обратную функцию к функции риска вида $1 - R(t|x)$.

Так как цель метода найти лучшую прогностическую модель среди набора всех базовых моделей, которая при этом обеспечивает хорошую калибровку функции риска, то необходимо перейти к проблеме оптимизации, как к задаче максимизации производительности модели дискриминантного анализа при условии ограничений на калибровку данных. Если обозначить за \mathfrak{R} набор всех функций риска базовых моделей, за f – метрику качества дискриминантной модели, за g – точность калибровки, то задача сводится к поиску функции $R^* \in \mathfrak{R}$, которая решает следующую задачу максимизации:

$$\begin{aligned} \max_{R \in \mathfrak{R}} f(R), \\ g(R) \leq c, \end{aligned} \quad (2)$$

где c – некоторая допустимая ошибка прогнозирования.

Алгоритм метода временного квилтинга с учетом введенных обозначений имеет следующий вид.

1. Построение базовых моделей выживаемости. То есть формирование множества $M = \{CoxPH, CoxPHRidge, Weibull...\}$.

2. Применение метода одеяла выживаемости для формирования функции риска. Прогнозы риска на прошлых временных горизонтах $k-1$ переносятся на будущий период k . Сначала функция риска для каждой модели разбивается на временные горизонты $T = \{t_0 = 0, t_1, \dots, t_k\}$. Затем на каждом временном интервале, присваивается весовой коэффициент для каждого приращения функции базовой модели выживаемости, а затем суммируются взвешенные комбинации приращения функции с течением времени. Определим функцию приращения для модели $m \in M$ на времени $[a, b]$ при определенном наборе входных предикторов

$$i_m(a, b|x) = R_m(b|x)R_m(a|x), \quad (3)$$

где R_m функция риска для модели m . Из-за того, что R_m не убывает на отрезке $[a, b]$, то и i_m не убывает и не отрицательна на отрезке времени $[a, b]$.

Обозначим через \mathbf{w} – вектор весовых коэффициентов размерности $|M|$, где $\mathbf{w}[m] \in [0,1]$, а также $\sum_{m \in M} \mathbf{w}[m] = 1$. Учитывая \mathbf{w} , взвешенное приращение функции на интервале $[a, b]$ определяется как

$$I_w(a, b|x) = \sum_{m \in M} w[m] * i_m(a, b|x). \quad (4)$$

Затем, учитывая веса $w_1 \dots w_k$ и время $t = \{t_0, t_1, \dots, t_k\}$, получаем

$$R_0(t|x) = \sum_{l=1}^{k-1} I_{w_l}(t_{l-1}, t_l|x) + I_{w_k}(t_{k-1}, t|x), \quad (5)$$

где первое слагаемое – это совокупный риск до времени t_{k-1} , а второе слагаемое – приращение от времени t_{k-1} до момента $t \in [t_{k-1}, t_k]$. Тогда функцию риска в момент времени t можно выразить следующим образом:

$$R(t|x) = \min \{1, R_0(t|x)\} \quad (6)$$

Обратим внимание, что $R_0(0|x) = 0$, то есть пациенты живы на начало периода наблюдения. На практике R_0 может превышать 1, и в этом случае он не может быть допустимой функцией риска. Этого можно избежать, установив $R = \min \{1, R_0\}$.

3. Построение модели байесовской оптимизации с ограничениями, которая моделирует показатели точности детерминации и калибровки в виде функций черного ящика, входными данными которых является массив весов (за разные временные горизонты), а выходными данными являются соответствующие достигнутые показатели.

Предположим, что конфигурация весов $W_k = w_1 \dots w_k$ – это шаблон квилтинга. Напомним, что цель – найти наилучшую функцию риска, определенную формулой (6). Поэтому задача сводится к поиску лучшего набора весов W_k^* , то есть набора весов, который решит задачу оптимизации

$$\begin{aligned} \max_{W_k} f(W_k), \\ g(W_k) \leq c, \end{aligned} \quad (7)$$

где $c > 0$ – некоторая допустимая ошибка прогнозирования.

В формуле (7) в качестве функции f принимается среднее значений функций f_k . Аналогично функция g – это среднее от функций g_k , являющихся метрикой характеристики, зависящей от времени калибровки.

Поэтому, формально $f(W_k)$ и $g(W_k)$ можно выразить как

$$f(W_k) = \frac{1}{K} \sum_{k=1}^K f_k(W_k) \quad (8)$$

$$g(W_k) = \frac{1}{K} \sum_{k=1}^K g_k(W_k). \quad (9)$$

Поскольку, в формуле (7) целевая функция и функция ограничения имеют неаналитическую форму, то их можно рассматривать как «черный ящик» $f, g: [0,1]^K \rightarrow R$.

Для поиска оптимального шаблона для квилтинга W_k^* используется байесовская оптимизация. Алгоритм байесовской оптимизации определяет гауссовский процесс (GP) на функциях f, g

$$f \sim GP(\mu_f(W_k), \kappa_f(W_k, W_k')) \quad (12)$$

$$g \sim GP(\mu_g(W_k), \kappa_g(W_k, W_k')), \quad (13)$$

где $\mu_f(W_k)$ и $\mu_g(W_k)$ – это средние функции, кодирующие ожидаемую производительность различных шаблонов квилтинга, а $\kappa_f(W_k, W_k')$ и $\kappa_g(W_k, W_k')$ – ядра

предикторов, кодирующие сходство между разными шаблонами квилтинга для f и g соответственно. В [11] эта проблема оптимизации называется проблемой композиции лоскутного шитья (Quilting Pattern Composition Problem (QPCP)).

4. Решение задачи байесовской оптимизации с ограничениями. На шаге k , чтобы решить задачу с ограничениями «черного ящика», в (7) используем расширенную функцию Лагранжа.

$$L(W_k, \lambda, \rho) = -f_k(W_k) + \lambda(g_k(W_k) - c) + \frac{1}{\rho} \max(0, g_k(W_k) - c)^2, \quad (10)$$

где $\rho > 0$ и $\lambda \geq 0$ обозначают параметр штрафа и множитель Лагранжа соответственно.

5. Формирование окончательной модели на основе результирующего набора весов.

Схематически алгоритм приведен на Рисунке 2.

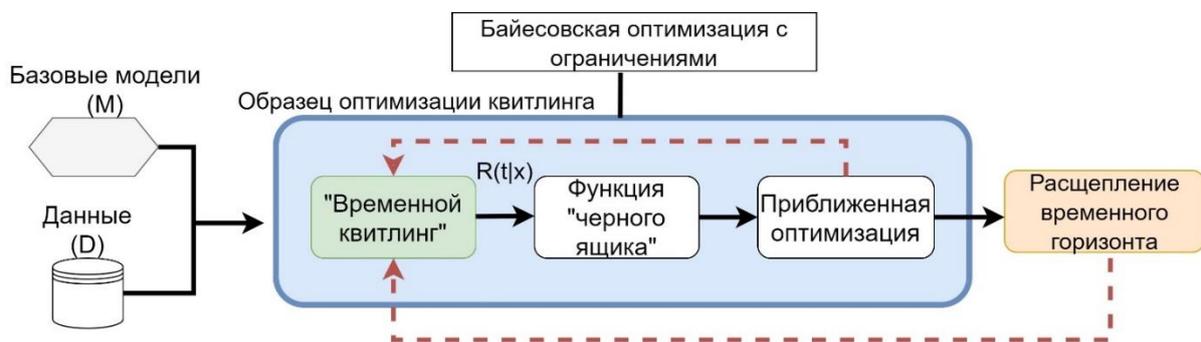


Рисунок 2 – Схематическое изображение метода «одеяло выживания» и его шаблона оптимизации на шаге k

Figure 2 – A schematic depiction of Survival Quilts and its pattern optimization at step k

Метрики оценки производительности

Оценка базовых моделей производилась несколькими способами:

1. Оценка риска с помощью встроенной функции библиотеки `scikit-survival` – `predict_survival_function`.
2. Оценка точности модели с помощью метода `score()` как на тренировочном, так и на тестовом множествах.
3. Оценка модели с помощью матрицы классификации: доля правильных ответов, чувствительность и специфичность.
4. Расчет ошибки калибровки с помощью пользовательской функции.

Методы 1-3 ранее упоминались в исследовании [17]. Рассмотрим подробнее метод 4. Когда предсказанные вероятности хорошо соответствуют реальным вероятностям, они называются откалиброванными. Для оценки модели необходим реальный количественный способ измерения точности калибровки. Наиболее часто используемая метрика называется ожидаемой ошибкой калибровки. Данная метрика показывает, насколько далеко в среднем наша спрогнозированная вероятность от настоящей вероятности. Чем меньше данный показатель – тем лучше прогностическая способность модели.

Легко определить погрешность калибровки одного временного промежутка: нужно взять модуль разности среднего значения прогнозируемых вероятностей и доли положительных результатов в рамках этого промежутка. Таким образом, ожидаемая ошибка калибровки (ECE) представляет собой средневзвешенное значение ошибок

калибровки отдельных периодов, где каждая период весит пропорционально количеству наблюдений, которые он содержит:

$$ECE = \frac{\sum_{b=1}^B |mean(y_b - mean(proba_b))| * len(y_b)}{\sum_{b=1}^B len(y_b)} \quad (11)$$

где b определяет период, а B – количество периодов. В знаменателе общее число примеров из выборки.

Чтобы определить количество периодов B в формуле (11) используется правило Фридмана-Диакониса (статистическое правило, предназначенное для определения количества разбиений, которое делает гистограмму максимально близкой к теоретическому распределению вероятностей).

Результаты и их обсуждение

При построении моделей машинного обучения исходная выборка разбиралась на тестовое и обучающее множества в соотношении 20:80. Оценка качества модели производилась по метрикам, описанным выше. В качестве выходной переменной использовалось бинарное значение – выживет ли пациент после инфаркта (1 – умрет, 0 – выживет), в качестве входных параметров использовались исходные клинические показатели [16]. Стандартные метрики качества модели рассчитаны с учетом прогноза выживаемости на 20 день. Данное значение выбрано на основе результатов исследований, проведенных ранее [10].

В Таблице 1 приведены результаты точности моделей на тестовые и обучающие множества. Можно отметить, что все модели имеют достаточно высокую точность.

Таблица 1 – Точность моделей на тестовом и обучающем множествах

Table 1 – Accuracy of models on test and training datasets

Метод	Точность на обучающ. мн-ве	Точность на тестовом мн-ве
CoxPH	0.821	0.821
CoxPHRidge	0.821	0.821
Weibull	0.821	0.821
LogLogistic	0.823	0.822
LogNormal	0.831	0.822
RandomSurvForest	0.837	0.791

В Таблице 2 отображены основные метрики качества моделей на тестовом множестве. Помимо них приведены значения ошибки калибровки, полученные пользовательской функцией. Хуже всего откалибрована модель случайного леса – RandomSurvForest. Однако данная модель показывает самую высокую точность, что важно для задач медицинского характера [17]. Модель временного квилтинга оказалась не самой откалиброванной, хотя метрики качества говорят о том, что модель достаточно эффективная. Она показывает высокую точность модели, высокую специфичность и достаточно низкую ошибку калибровки.

Помимо количественной оценки, хороший способ оценить калибровку модели – построить график, называемый «калибровочной кривой» или «диаграммой надежности». Идея состоит в том, чтобы разделить наблюдения на группы таким образом, чтобы наблюдения, принадлежащие к одной и той же группе, имели одинаковую вероятность. Для каждой группы наблюдений калибровочная кривая сравнивает спрогнозированное среднее (т. е. среднее значение предсказанной вероятности) с теоретическим средним (т. е. средним наблюдаемой целевой переменной).

График строится с помощью функции `calibration_curve`, библиотеки Scikit-learn. Необходимо выбрать количество групп, и можно указать необязательный параметр – стратегия распределения. Существует две вариации стратегии распределения для данной функции: «uniform» – интервал от 0 до 1, разделенный на количество отрезков одинаковой длины; «quantile» – границы разбиения вероятностей распределяются таким образом, чтобы в каждой группе было одинаковое количество наблюдений.

Таблица 2 – Метрики качества моделей

Table 2 – Model quality metrics

Метод	Доля прав.отв.	Чувствительность	Специфичн.	Ошибка калибр.
CoxPH	0.838	0.403	0.939	0.062
CoxPHRidge	0.839	0.405	0.939	0.061
Weibull	0.832	0.436	0.923	0.043
LogLogistic	0.822	0.462	0.904	0.023
LogNormal	0.825	0.448	0.913	0.032
RandomSurvF.	0.849	0.339	0.967	0.094
Survival Quilts	0.846	0.41	0.97	0.04

Если у модели хорошая точность, тогда калибровочная кривая будет монотонно увеличиваться. Но это не значит, что модель хорошо откалибрована. Модель хорошо откалибрована только в том случае, если калибровочная кривая очень близка к биссектрисе, поскольку это будет означать, что прогнозируемая вероятность в среднем близка к теоретической вероятности. Существует четыре типа графиков, указывающих на отклонение калибровки:

«Систематическая переоценка». По сравнению с правильным распределением распределение прогнозируемых вероятностей смещается вправо. Так часто случается, когда модель обучается на несбалансированном наборе данных с очень небольшим количеством наблюдений в одном из классов.

«Систематическая недооценка». По сравнению с правильным распределением распределение прогнозируемых вероятностей смещается влево.

«Центр распределения слишком тяжелый». Так происходит, когда модель имеет тенденцию отклонять прогнозируемые вероятности от 0 и 1.

«Хвосты распределения слишком тяжелые». Так происходит, когда модель имеет тенденцию приближать прогнозируемые вероятности от 0 и 1.

На Рисунке 3 показаны калибровочные кривые для базовых моделей текущего исследования. Непараметрические модели пропорциональных рисков Кокса и все параметрические модели относятся к типу систематической переоценки. Для модели случайного леса центр распределения слишком тяжелый. На Рисунке 4 отдельно отображена калибровочная кривая для метода временного квилтинга. Данный график тоже можно отнести к типу «центр распределения слишком тяжелый», однако данная кривая наиболее приближена к биссектрисе, что говорит о лучшем распределении вероятностей модели.

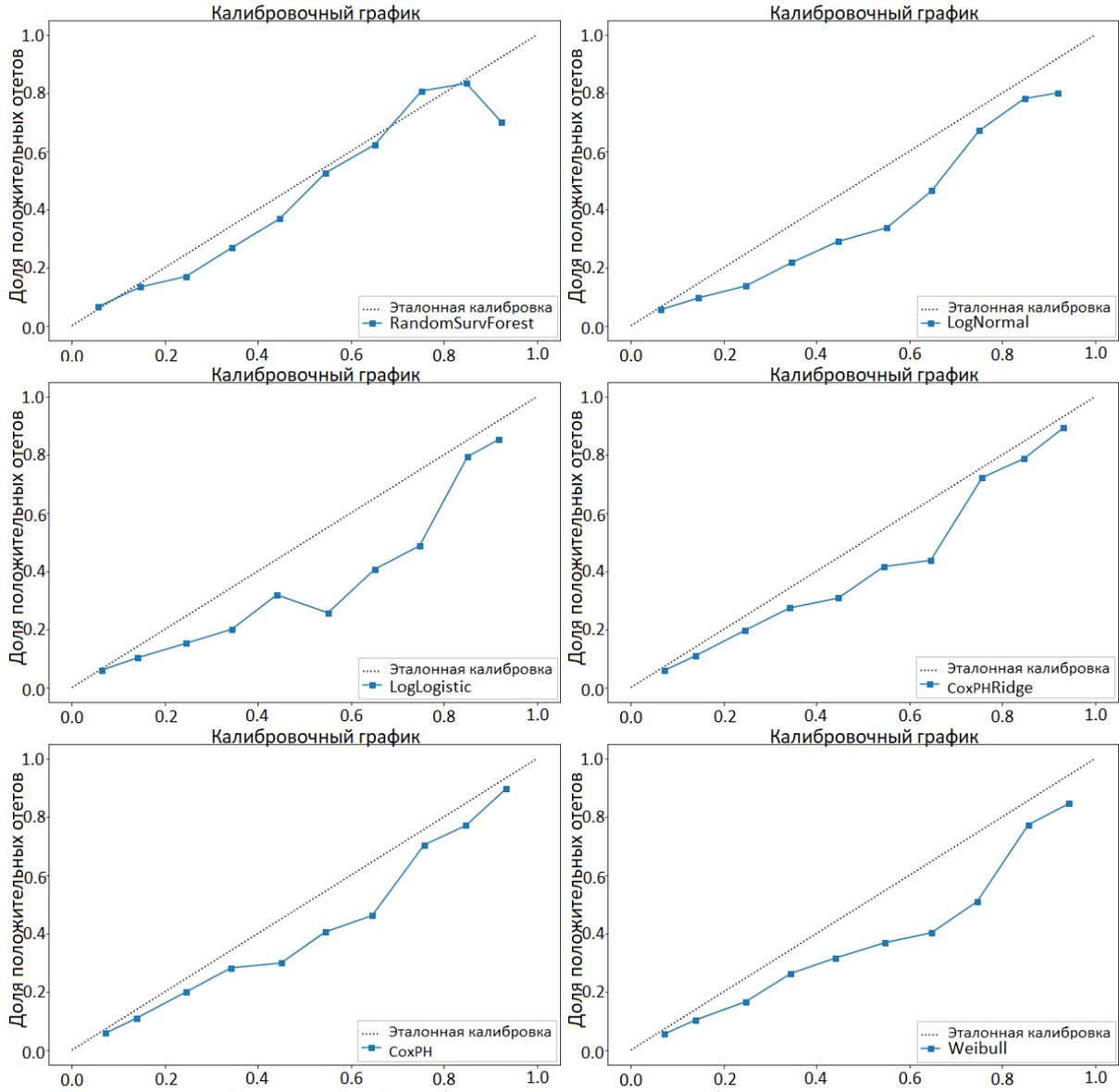


Рисунок 3 – Калибровочные кривые для базовых моделей
Figure 3 – Calibration Curves for Basic Models

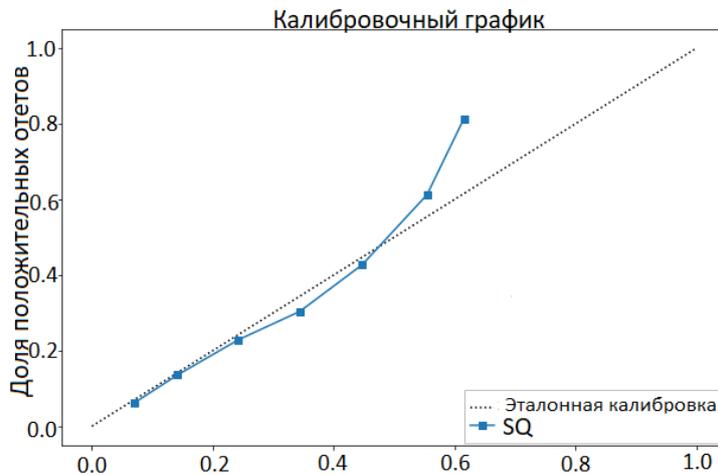


Рисунок 4 – Калибровочная кривая для модели временного квилтинга
Figure 4 – Calibration Curve for Temporal Quilting model

Заключение

В ходе данного исследования было рассмотрено несколько моделей анализа выживаемости после инфаркта миокарда: непараметрические методы (модель пропорциональных рисков Кокса, модель пропорциональных рисков Кокса с использованием гребневой регрессии), параметрические модели (логистическая модель нормального распределения, логистическая модель экспоненциального распределения, метода распределения Вейбулла) и ансамблевая модель (случайный лес). Также был применен усовершенствованный метод – временной квилтинг, который создает изменяющиеся во времени ансамбли существующих моделей выживания. Данная модель продемонстрировала наилучшие показатели оценки калибровки модели, при тех же показателях точности. Также была проанализирована метрика ошибки калибровки модели.

По результатам исследования наиболее эффективной оказалась модель временного квилтинга, наименее эффективной – модель случайного леса.

СПИСОК ИСТОЧНИКОВ

1. Greg Ridgeway. The state of boosting. *Computing Science and Statistics*. 1999; 31:172–181.
2. Hosmer DW, Lemeshow S, May S. *Applied survival analysis regression modeling of time-to-event data, 2nd ed*. Hoboken, NJ: Wiley-Interscience; 2008. 2006 p.
3. Austin P. Generating survival times to simulate cox proportional hazards models with time-varying covariates. *Statistics in medicine*. 2012; 31(29):3946–3958. DOI: 10.1002/sim.5452
4. Katzman J., Shaham U., Bates J. Deep survival: A deep cox proportional hazards network. *BMC Medical Research Methodology*. 2016;18(24):1–15. DOI: 10.1186/s12874-018-0482-1
5. Ahmed M., Mihaela van der Schaar. Deep multi-task gaussian processes for survival analysis with competing risks. *In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*. Curran Associates Inc., Red Hook, NY, USA;2017: 2326–2334.
6. Bellot A., Mihaela van der Schaar. Boosted trees for risk prognosis. *In Proceedings of the 3st Machine Learning for Healthcare Conference (MLHC 2018)*. 2018; PMLR (85):2–16.
7. Taser PY. Application of Bagging and Boosting Approaches Using Decision Tree-Based Algorithms in Diabetes Risk Prediction. *Proceedings*. 2021;74(1):6. DOI: 10.3390/proceedings2021074006
8. Lee C, Zame W, Yoon J, van der Schaar M. DeepHit: A Deep Learning Approach to Survival Analysis With Competing Risks. 2018; 32(1). Режим доступа: <https://ojs.aaai.org/index.php/AAAI/article/view/11842> (дата обращения: 01.10.2021).
9. Spooner A., Chen E., Sowmya A. A comparison of machine learning methods for survival analysis of high-dimensional clinical data for dementia prediction. *Sci Rep* 2020;10:20410. DOI: 10.1038/s41598-020-77220-w
10. Firyulina M., Bondarenko Yu., Desyatirikova E. Identification of Risk Factors for Mortality after Myocardial Infarction Using Machine Learning Methods. *Proc. of 2021 24th International Conference on Soft Computing and Measurements*. SCM. 2021. DOI: 10.1109/SCM52931.2021.9507190
11. Lee C., Zame W., Alaa A. Temporal Quilting for Survival Analysis. *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics in Proceedings of Machine Learning Research, PMLR*. 2019; 89:596–605.

12. Naeini MP, Cooper GF, Hauskrecht M. Obtaining Well Calibrated Probabilities Using Bayesian Binning. *Proc Conf AAAI Artif Intell.* 2015; 2015:2901–2907.
13. Guo C., Pleiss G., Kilian Q. On calibration of modern neural networks. *In Proceedings of the 34th International Conference on Machine Learning.* Weinberger. 2017:1321–1330.
14. Niculescu-Mizil, Alexandru & Caruana, Rich. (2005). Predicting good probabilities with supervised learning. *ICML 2005 - Proceedings of the 22nd International Conference on Machine Learning.* 2005:625–632. DOI:10.1145/1102351.1102430.
15. Merkle, Edgar & Hartman, R. Weighted Brier score decompositions for topically heterogenous forecasting tournaments. *Judgment and Decision Making.* 2018;13(2):185–201.
16. Firyulina M.A., Kashirina I.L. Classification of cardiac arrhythmia using machine learning techniques. *Journal of physics: Applied Mathematics, Computational Science and Mechanics: Current Problems.* 2019:167–1175.
17. Kashirina I., Firyulina M. Building models for predicting mortality after myocardial infarction in conditions of unbalanced classes, including the influence of weather conditions. *CEUR Workshop Proceedings.* 2020;2790:188–197
18. Jasper S., Larochelle H., Adams R. *Practical Bayesian Optimization of Machine Learning Algorithms.* Curran Associates, Inc.; 2012. 25 p.
19. Feurer M., Hutter F. *Automated Machine Learning.* Cham, The Springer Series on Challenges in Machine Learning; 2019. 223 p.
20. Hutter F., Hoos H., Leyton-Brown K. Sequential Model-Based Optimization for General Algorithm Configuration. *Lecture Notes in Computer Science.* 2011;6683:507–523. DOI:10.1007/978-3-642-25566-3_40
21. Thornton C., Hutter F., Hoos H. Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms. *Knowledge Discovery and Data Mining.* 2013;6683:847–855. DOI:10.1145/2487575.2487629

REFERENCES

1. Greg Ridgeway. The state of boosting. *Computing Science and Statistics.* 1999;31:172–181.
2. Hosmer DW, Lemeshow S, May S. *Applied survival analysis regression modeling of time-to-event data, 2nd ed.* Hoboken, NJ: Wiley-Interscience; 2008. 2006 p.
3. Austin P. Generating survival times to simulate cox proportional hazards models with time-varying covariates. *Statistics in medicine.* 2012; 31(29):3946–3958. DOI: 10.1002/sim.5452
4. Katzman J., Shaham U., Bates J. Deep survival: A deep cox proportional hazards network. *BMC Medical Research Methodology.* 2016;18(24):1–15. DOI: 10.1186/s12874-018-0482-1
5. Ahmed M., Mihaela van der Schaar. Deep multi-task gaussian processes for survival analysis with competing risks. *In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17).* Curran Associates Inc., Red Hook, NY, USA; 2017:2326–2334.
6. Bellot A., Mihaela van der Schaar. Boosted trees for risk prognosis. *In Proceedings of the 3st Machine Learning for Healthcare Conference (MLHC 2018).* 2018; PMLR (85):2–16.
7. Taser PY. Application of Bagging and Boosting Approaches Using Decision Tree-Based Algorithms in Diabetes Risk Prediction. *Proceedings.* 2021;74(1):6. DOI: 10.3390/proceedings2021074006

8. Lee C, Zame W, Yoon J, van der Schaar M. DeepHit: A Deep Learning Approach to Survival Analysis With Competing Risks. 2018;32(1). Available from: <https://ojs.aaai.org/index.php/AAAI/article/view/11842> (accessed 01.10.2021)
9. Spooner A., Chen E., Sowmya A. A comparison of machine learning methods for survival analysis of high-dimensional clinical data for dementia prediction. *Sci Rep* 2020;10:20410. DOI: 10.1038/s41598-020-77220-w
10. Firyulina M., Bondarenko Yu., Desyatirikova E. Identification of Risk Factors for Mortality after Myocardial Infarction Using Machine Learning Methods. *Proc. of 2021 24th International Conference on Soft Computing and Measurements*. SCM. 2021. DOI: 10.1109/SCM52931.2021.9507190
11. Lee C., Zame W., Alaa A. Temporal Quilting for Survival Analysis. *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics in Proceedings of Machine Learning Research, PMLR*. 2019; 89:596–605.
12. Naeini MP, Cooper GF, Hauskrecht M. Obtaining Well Calibrated Probabilities Using Bayesian Binning. *Proc Conf AAAI Artif Intell*. 2015;2015:2901–2907.
13. Guo C., Pleiss G., Kilian Q. On calibration of modern neural networks. *In Proceedings of the 34th International Conference on Machine Learning*. Weinberger. 2017:1321–1330.
14. Niculescu-Mizil, Alexandru & Caruana, Rich. (2005). Predicting good probabilities with supervised learning. *ICML 2005 - Proceedings of the 22nd International Conference on Machine Learning*. 2005:625–632. DOI:10.1145/1102351.1102430.
15. Merkle, Edgar & Hartman, R. Weighted Brier score decompositions for topically heterogenous forecasting tournaments. *Judgment and Decision Making*. 2018;13(2):185–201.
16. Firyulina M.A., Kashirina I.L. Classification of cardiac arrhythmia using machine learning techniques. *Journal of physics: Applied Mathematics, Computational Science and Mechanics: Current Problems*. 2019:167–1175.
17. Kashirina I., Firyulina M. Building models for predicting mortality after myocardial infarction in conditions of unbalanced classes, including the influence of weather conditions. *CEUR Workshop Proceedings*. 2020;2790:188–197
18. Jasper S., Larochelle H., Adams R. *Practical Bayesian Optimization of Machine Learning Algorithms*. Curran Associates, Inc.; 2012. 25 p.
19. Feurer M., Hutter F. *Automated Machine Learning*. Cham, The Springer Series on Challenges in Machine Learning; 2019. 223 p.
20. Hutter F., Hoos H., Leyton-Brown K. Sequential Model-Based Optimization for General Algorithm Configuration. *Lecture Notes in Computer Science*. 2011;6683:507–523. DOI:10.1007/978-3-642-25566-3_40
21. Thornton C., Hutter F., Hoos H. Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms. *Knowledge Discovery and Data Mining*. 2013; 6683:847–855. DOI:10.1145/2487575.2487629

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Фирюлина Мария Андреевна, аспирант,
кафедра математических методов исследования
операций, ФГБОУ ВО «Воронежский
государственный университет», Воронеж,
Российская Федерация.
e-mail: mashafiryulina@mail.ru
ORCID: [0000-0003-3468-5514](https://orcid.org/0000-0003-3468-5514)

Mariya A. Firyulina, Phd Student,
Mathematical Methods of Operations
Research Department, Voronezh state
university, Voronezh, Russian Federation.

*Статья поступила в редакцию 10.11.2021; одобрена после рецензирования 22.12.2021;
принята к публикации 27.12.2021.*

*The article was submitted 10.11.2021; approved after reviewing 22.12.2021;
accepted for publication 27.12.2021.*