МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ, ЧИСЛЕННЫЕ МЕТОДЫ И КОМПЛЕКСЫ ПРОГРАММ/MATHEMATICAL MODELING, NUMERICAL METHODS AND PROGRAM COMPLEXES

DOI: https://doi.org/10.60797/IRJ.2025.162.28

РЕАЛИЗАЦИЯ АЛГОРИТМА МУРАВЬЕВ-ОПЫЛИТЕЛЕЙ С МАТРИЧНОЙ СТРУКТУРОЙ ЦЕЛЕВОЙ ФУНКЦИИ ОПТИМИЗАЦИИ

Научная статья

Микулик И.И.[1], *, Благовещенская Е.А.[2], Аристархов Ф.Е.[3], Фесак Д.И.[4]
[1] ORCID : 0000-0002-0542-7914;
[2] ORCID : 0000-0002-2425-5556;
[1, 2, 3, 4] Петербургский государственный университет путей сообщения Императора Александра I, Санкт-Петербург, Российская Федерация

* Корреспондирующий автор (guess_97[at]mail.ru)

Аннотация

Статья посвящена разработке метода построения интерпретируемых моделей анализа выживаемости. В качестве основы используется расширенная модель Кокса, где зависимость между признаками задается полиномом. Для решения задачи оптимизации структуры полинома предложена модификация алгоритма муравьев-опылителей. Ключевой особенностью является матричное представление целевой функции, объединяющей критерии точности (c-индекс), количество признаков и сложность модели. В отличие от классических подходов, феромон в алгоритме откладывается на вершинах графа признаков. Метод был протестирован на данных о критических неисправностях 5000 автомобилей. В ходе эксперимента алгоритм продемонстрировал эффективность, точно восстанавливая заданную зависимость функции риска от признаков при среднем количестве итераций 6,09. Результаты подтверждают, что предложенный подход позволяет одновременно строить точные прогностические модели и выполнять отбор значимых признаков, обеспечивая высокую интерпретируемость результатов. Перспективы работы связаны с развитием матричного представления феромона и апробацией на реальных данных.

Ключевые слова: анализ выживаемости, метод муравьев-опылителей, методы оптимизации, матричная целевая функция.

IMPLEMENTATION OF THE ANT POLLINATOR ALGORITHM WITH A MATRIX STRUCTURE OF THE OPTIMISATION GOAL FUNCTION

Research article

Mikulik I.I.[1], *, Blagoveshchenskaya Y.A.[2], Aristarkhov F.Y.[3], Fesak D.I.[4]
[1] ORCID : 0000-0002-0542-7914;
[2] ORCID : 0000-0002-2425-5556;
[1, 2, 3, 4] Emperor Alexander I St. Petersburg State Transport University, Saint-Petersburg, Russian Federation

* Corresponding author (guess_97[at]mail.ru)

Abstract

The article is devoted to the development of a method for constructing interpretable survival analysis models. An extended Cox model is used as a basis, where the dependence between traits is specified by a polynome. To solve the problem of optimising the polynomial structure, a modification of the ant-pollinator algorithm is suggested. A key feature is the matrix representation of the goal function, which combines accuracy criteria (c-index), the number of features, and model complexity. Unlike classical approaches, the pheromone in the algorithm is deposited at the vertices of the trait graph. The method was tested on data about critical malfunctions in 5,000 cars. During the experiment, the algorithm demonstrated its effectiveness, accurately restoring the specified dependence of the risk function on features with an average number of iterations of 6.09. The results confirm that the proposed approach allows to simultaneously build accurate predictive models and select significant features, ensuring high interpretability of the results. The prospects for further work are related to the development of a matrix representation of pheromones and testing on real data.

Keywords: survival analysis, ant pollinator method, optimisation methods, matrix goal function.

Введение

Прогнозирование наступления терминальных событий, таких как отказ технических систем, является критически важной задачей в инженерии, анализе надежности, а также в медицине [1]. Широко распространенная модель Кокса [2], несмотря на свою интерпретируемость, опирается на предположение о линейной зависимости между признаками и функцией риска, что часто нарушается на практике [3]. Хотя методы машинного обучения способны определять нелинейные зависимости, их результаты часто представляют собой «черный ящик», что затрудняет их практическое использование специалистами. В связи с этим актуальной является разработка методов, сочетающих высокую адаптивность к сложным данным и прозрачность получаемых моделей.

Объектом исследования является расширенная модель Кокса. Предметом исследования — метод построения расширенной модели Кокса. Цель работы заключается в возможности построения точной и интерпретируемой модели анализа выживаемости за счет разработки модификации алгоритма муравьев-опылителей, оперирующего матричным представлением целевой функции.

Научная новизна заключается в следующем. Предложена модификация алгоритма муравьев-опылителей, в которой целевая функция представлена в виде оптимизируемой матрицы, что позволяет эффективно решать многокритериальную задачу построения модели. Разработанный метод позволяет одновременно решать задачу построения точной прогностической модели и отбора наиболее значимых признаков. Для проверки эффективности предложенного метода был использован синтетический набор данных, содержащий параметры эксплуатации и время до выхода из строя для 5000 автомобилей. Использование искусственно сгенерированных данных с заранее известной зависимостью позволило объективно оценить способность алгоритма к воспроизведению истинных функциональных связей.

Методы и принципы исследования

2.1. Расширенная модель Кокса

Широко используемой моделью анализа выживаемости [4] является регрессионная модель пропорциональных рисков Кокса. Одним из достоинств модели является то, что функция риска этой модели содержит две компоненты — параметрическую, зависящую от параметров, и не параметрическую λ_0, называемую базовым риском. Поэтому модель Кокса называют полупараметрической моделью. Функция риска модели Кокса описывается в следующем виде:

$$\lambda(t, \mathbf{x}) = \lambda_0(t) \exp(\beta \cdot \mathbf{x}) \qquad (1)$$

где $\mathbf{x}$ — вектор значений признаков объекта данных, β — вектор влияния признаков данных.

Регрессионная модель Кокса, включающая параметрический и непараметрический компоненты, позволяет получить более согласованные оценки в широком диапазоне условий по сравнению с параметрическими моделями и более точные оценки по сравнению с непараметрическими методами [5].

Использовав иную, нелинейную зависимость g между признаками, можно построить класс моделей, расширяющих классическую модель Кокса:

$$\lambda(t, \mathbf{x}) = \lambda_0(t) \exp(g(\beta, \mathbf{x})) \qquad (2)$$

Функция (2) является обобщением функции (1).

В данной работе такой функцией является полином специального вида: он является суммой одночленов, каждый из которых является произведением признаков, при том степень каждого признака в одночлене не превосходит единицы.

2.2. Постановка задачи

Признанным способом оценки точности моделей анализа выживаемости является c-индекс (индекс согласованности) [6]. В его основе лежит ранговая корреляция между фактическим временем выживания и прогнозами модели. Он представляет точность ранжирования прогнозов модели выживаемости, показывая, насколько хорошо модель упорядочивает объекты по вероятности выживания [7]. Является величиной, принимающей значения от нуля до единицы, где 1 — идеальное соответствие, 0.5 — соответствие уровню случайного угадывания.

Задачей является построение модели по набору данных. Так как моделью в данном исследовании является расширенная модель Кокса, необходимо построить полином, c-индекс которого, был бы наибольшим, а количество признаков — минимальным. Также для повышения интерпретируемости модели и для борьбы с ошибкой переобучения имеет смысл ввести показатели сложности построенного полинома, которые также необходимо минимизировать.

2.3. Матричное представление целевой функции

Исходя из постановки задачи, введены четыре критерия, которые будут оптимизированы.

Первый критерий — точность модели c, которая оценивается с помощью c-индекса. Второй критерий отвечает за количество признаков, необходимых для обучения модели (2) $r = \frac{|F|}{p}$, где $|F|$ — количество признаков найденного решения, а p — количество всех возможных признаков. Следующий критерий k отвечает за количество аддитивных связей построенного полинома: $k = \frac{\log_2 K}{p}$, где K — количество мономов построенного полинома. Последним критерий отвечает за количество мультипликативных связей $b = \frac{K+B}{Kp}$, где B — количество мультипликативных связей в построенном полиноме.

Таким образом, с учетом введенных критериев задача оптимизация заключается в поиске полинома при условиях

$$\begin{cases} \mathrm{c} \to \max \\ r \to \min \\ k \to \min \\ b \to \min \end{cases} \qquad (3)$$

Система условий (3) порождает многокритериальную задачу оптимизации. Для ее решения предложено рассмотрение условий в виде матрицы:

$$\begin{pmatrix} c & b \\ k & 1 - r \end{pmatrix} \qquad (4)$$

Тогда определитель матрицы (4) сводит задачу к одномерной оптимизации.

2.4. Метод муравьев-опылителей

Решение задачи оптимизации осуществляется с помощью метаэвристического гибридного метода муравьев-опылителей [8]. Алгоритм состоит из трех этапов: муравьиный, генетический и этап опыления. Основное структурное отличие от классических муравьиных алгоритмов заключается в том, что процесс оптимизации происходит не на полном графе путей, а на графе признаков, где феромон откладывается на вершинах, а не на дугах. Это позволяет алгоритму оценивать и отбирать наиболее важные с точки зрения признаки и их взаимодействия.

Основные понятия алгоритма:

1. Цветок (вершина графа) — элементарная единица модели, соответствующая моному полинома. Исходное множество вершин инициализируется множеством признаков объекта.

2. Путь муравья — упорядоченный набор вершин (мономов), формирующий полиномиальную целевую функцию.

3. Феромон — числовая мера, связанная с каждой вершиной, которая динамически обновляется и отражает важность признака в построенных решениях.

4. Мера пути: Значение целевой функции, вычисленное для модели с полиномом, соответствующему пройденному пути. Значением является определитель матрицы (4).

Алгоритм является итерационным и состоит из трех последовательно выполняемых этапов.

2.4.1. Муравьиный этап

На данном этапе популяция агентов-«муравьев» осуществляет поиск оптимального пути — набора вершин-мономов, формирующего полином. Каждый муравей строит путь, стохастически выбирая вершины для добавления в свой полином. Вероятность выбора вершины зависит от уровня феромона, связанного с ней. После оценки целевой функции для всех построенных полиномов (путей) происходит обновление феромона на вершинах. Вершины, входящие в лучшие решения, получают прирост феромона. Одновременно моделируется процесс его испарения на всех вершинах для избегания преждевременной сходимости алгоритма.

2.4.2. Генетический этап

Этап используется для оптимизации ключевых параметров муравьиного алгоритма для повышения эффективности поиска решения. Оптимизируются параметры: чувствительность муравьев к феромону, чувствительность муравьев к эвристической информации, количество откладываемого феромона. Особи в генетическом алгоритме представляют собой векторы параметров муравьиного алгоритма. Над популяцией особей применяются стандартные генетические операторы: выбора (селекция на основе приспособленности, определяемой качеством работы муравьиного этапа с данными параметрами), кроссинговера (рекомбинация параметров) и мутации (случайное возмущение значений параметров).

2.4.3. Этап опыления

Данный этап отвечает за эволюцию самого множества вершин-мономов, обеспечивая адаптацию модели к данным. Этап реализует динамическое обновление множества цветков путем добавления новых перспективных мономов и удаления неэффективных. Эффективность вершины-монома определяется накопленным на ней уровнем феромона. Этап представляет собой последовательное применение операторов: селекции (отбор наименее эффективных вершин для удаления), кроссбридинга (умножение мономов), лайнбридинга (случайное добавление монома, состоящего из одного признака), старения (механизм планомерного уменьшения феромона на вершинах, который позволяет постепенно исключать устаревшие или неперспективные мономы из рассмотрения)

Совместная работа трех этапов образует метаэвристический механизм, который формирует структуру полиномиальной модели, обеспечивая как высокую точность, так и интерпретируемость результата.

Основные результаты

Метод муравьев-опылителей реализован в виде программы, написанной на языке программирования Python3. Использована библиотека Lifelines, предназначенная для работы с моделями и методами анализа выживаемости [9]. Она предоставляет интерфейс и реализацию основных моделей анализа выживаемости, в том числе модель пропорциональных рисков Кокса. Одним из преимуществ интерфейса является возможность создание собственных моделей и конструирование функции риска [10].

Для проведения вычислительного эксперимента сгенерирован набор данных о неисправностях автомобилей. Искусственная генерация данных позволила задать аналитическую форму зависимости времени до наступления терминального события от признаков. Это дало возможность оценить способность алгоритма к воспроизведению истинных функциональных зависимостей, скрытых в данных. Сгенерированный набор данных включает параметры эксплуатации и время выхода из строя для 5000 автомобилей. Набор данных содержит 10 признаков и время эксплуатации. Полином, использующийся для генерации данных: $age \cdot mileage$

Таблица 1 - Описание признаков тестируемого набора данных

DOI: https://doi.org/10.60797/IRJ.2025.162.28.1

Признак	Полное название	Значение
mileage	Пробег автомобиля	Числовой
age	Возраст автомобиля	Числовой
offroad_ratio	Доля поездок по бездорожью	Числовой
maintenance_reg	Техническое обслуживание	0 — присутствует 1 — отсутствует
engine_temp	Температура двигателя	Числовой

Признак	Полное название	Значение
tire_pressure	Давление в шинах	Числовой
battery_health	Состояние аккумулятора	Числовой
color	Цвет автомобиля	0 — белый 1 — серый 2 — черный 3 — другой
fuel_type	Вид топлива	0 — бензин 1 — дизель 2 — газ

Таким образом, для синтетического набора данных значимыми являются 2 признака из 10.

В проведенном эксперименте использован следующий набор параметров:

1. Исходная чувствительность муравьев к феромону $\alpha_0 = 0,5$.
2. Исходная чувствительность муравьев к эвристической информации $\beta_0 = 2$.
3. Исходное количество откладываемого феромона $Q_0 = 25$.
4. Испарение $\rho_0 = 0,8$.
5. Количество муравьев $n = 12$.
6. Вероятность кроссинговера $p_{cross} = 0,9$
7. Вероятность мутации $p_{mut} = 0,2$

Вычислительный эксперимент состоял из 100 запусков алгоритма. Во всех случаях алгоритм успешно идентифицировал полином, соответствующий функции $age \cdot mileage$, использовавшейся для генерации данных. Среднее число итераций, потребовавшееся для нахождения точного полинома, составило 6,09.

Обсуждение

Проведенное исследование демонстрирует, что предложенная модификация алгоритма муравьев-опылителей является работоспособным и эффективным инструментом для построения расширенных моделей Кокса. Высокая частота успешного нахождения целевого полинома и низкое среднее число итераций (6.09) свидетельствуют о хорошей сходимости метода и адекватности предложенного матричного подхода для решения многокритериальной задачи, позволяющего сводить критерии оптимизации в скалярную целевую функцию. Эти результаты подтверждают, что переход от оптимизации на графе путей к оптимизации на графе признаков с откладыванием феромона на вершинах является результативным для задач структурного синтеза моделей.

Основным преимуществом гибридного подхода является объединение его трех этапов. Муравьиный этап эффективно исследует пространство возможных полиномов, генетический алгоритм адаптивно настраивает параметры этого поиска, а этап опыления динамически обогащает и очищает словарь мономов, предотвращая стагнацию. Одновременная оптимизация структуры модели и ее параметров позволяет алгоритму успешно решать задачу отбора признаков и построения интерпретируемой нелинейной зависимости.

Вместе с тем, текущая реализация метода имеет области для потенциального улучшения. Использование определителя матрицы критериев, хотя и доказало свою эффективность, является по сути линейной сверткой, которая может нивелировать сложные взаимоотношения между отдельными целевыми функциями. Это ограничение открывает путь для дальнейших исследований, направленных на разработку более совершенных механизмов агрегирования, таких как матричное представление не только целевой функции, но и самой структуры феромона.

Заключение

В данной работе успешно решена задача разработки метода построения точной и интерпретируемой модели анализа выживаемости. Цель работы достигнута за счет модификации алгоритма муравьев-опылителей, ключевой особенностью которой является матричное представление целевой функции. Это представление позволило свести многокритериальную задачу максимизации точности и минимизации сложности модели к задаче однокритериальной оптимизации.

Метод решает не только задачу аппроксимации, но и проблему отбора признаков, что важно для повышения интерпретируемости итоговой модели. Практическая значимость подтверждена результатами вычислительного эксперимента на синтетических данных, где алгоритм стабильно и быстро находил заданную зависимость, демонстрируя свою робастность и эффективность.

Таким образом, работа вносит вклад в область интеллектуального анализа данных и анализа выживаемости, предлагая современный инструмент для построения прогностических моделей, сочетающих высокую адаптивность и прозрачность. Перспективы дальнейших исследований включают апробацию метода на реальных наборах данных, а также развитие предложенной матричной парадигмы, в частности, во внедрении матричного феромона и соответствующих матричных операций для его обновления, что может существенно повысить универсальность алгоритма.

Финансирование

Исследование выполнено за счет гранта Российского научного фонда (проект № 25-11-00348).

Funding

This work was funded by the Russian Science Foundation grant number 25-11-00348.

Конфликт интересов

Не указан.

Conflict of Interest

None declared.

Рецензия

Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.

Review

All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

Список литературы / References

1. Purnami S.W. Cox Model Survival Analysis to Evaluate Treatment of Electro-Capacitive Cancer Therapy (ECCT) For Cancer Patients. / S.W. Purnami, R.S. Putra, A.I. Edina et al. // Journal of Physics: Conference Series. — 2021. — № 1. — DOI: 10.1088/1742-6596/1863/1/012036
2. Cox D.R. Regression models and life-tables. / D.R. Cox // Journal of the Royal Statistical Society: Series B (Methodological). — 1972. — № 2. — P. 187–202.
3. Austin P.S. The Impact of Violation of the Proportional Hazards Assumption on the Calibration of the Cox Proportional Hazards Model. / P.S. Austin, D. Giardiello // Statistics in Medicine. — 2025. — № 44. — DOI: 10.1002/sim.70161
4. Prentice R.L. Introduction to Cox (1972) regression models and life-tables. / R.L. Prentice. // Breakthroughs in Statistics: Methodology and Distribution; — 2. — New York: Springer New York, 1992. — P. 519–526.
5. Powell J.L. Estimation of semiparametric model. / J.L. Powell // Handbook of econometrics. — 1994. — № 4.
6. Alabdallah A. The Concordance Index decomposition: A measure for a deeper understanding of survival prediction models. / A. Alabdallah, M. Ohlsson, S. Pashami et al. // Artificial Intelligence in Medicine. — 2024. — № 148. — DOI: 10.1016/j.artmed.2024.102781
7. Longato E. A practical perspective on the concordance index for the evaluation and selection of prognostic time-to-event models. / E. Longato, M. Vettoretti, B. Di Camillo // Journal of Biomedical Informatics. — 2020. — № 108. — DOI: 10.1016/j.jbi.2020.103496
8. Микулик И.И. Методика для решения задачи выбора признаков в регрессионной модели Кокса. / И.И. Микулик // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. — 2024. — № 3. — С. 85–94. — DOI: 10.24143/2072-9502-2024-3-85-94
9. Davidson-Pilon C. lifelines: survival analysis in Python. / C. Davidson-Pilon // Journal of Open Source Software. — 2019. — № 40. — DOI: 10.21105/joss.01317
10. Faghiri F. Cox proportional hazards model with Bayesian neural network for survival prediction. / F. Faghiri, A. Kohansal // Scientific Reports. — 2025. — № 1. — DOI: 10.1038/s41598-025-16993-4

Список литературы на английском языке / References in English

1. Purnami S.W. Cox Model Survival Analysis to Evaluate Treatment of Electro-Capacitive Cancer Therapy (ECCT) For Cancer Patients. / S.W. Purnami, R.S. Putra, A.I. Edina et al. // Journal of Physics: Conference Series. — 2021. — № 1. — DOI: 10.1088/1742-6596/1863/1/012036
2. Cox D.R. Regression models and life-tables. / D.R. Cox // Journal of the Royal Statistical Society: Series B (Methodological). — 1972. — № 2. — P. 187–202.
3. Austin P.S. The Impact of Violation of the Proportional Hazards Assumption on the Calibration of the Cox Proportional Hazards Model. / P.S. Austin, D. Giardiello // Statistics in Medicine. — 2025. — № 44. — DOI: 10.1002/sim.70161
4. Prentice R.L. Introduction to Cox (1972) regression models and life-tables. / R.L. Prentice. // Breakthroughs in Statistics: Methodology and Distribution; — 2. — New York: Springer New York, 1992. — P. 519–526.
5. Powell J.L. Estimation of semiparametric model. / J.L. Powell // Handbook of econometrics. — 1994. — № 4.
6. Alabdallah A. The Concordance Index decomposition: A measure for a deeper understanding of survival prediction models. / A. Alabdallah, M. Ohlsson, S. Pashami et al. // Artificial Intelligence in Medicine. — 2024. — № 148. — DOI: 10.1016/j.artmed.2024.102781
7. Longato E. A practical perspective on the concordance index for the evaluation and selection of prognostic time-to-event models. / E. Longato, M. Vettoretti, B. Di Camillo // Journal of Biomedical Informatics. — 2020. — № 108. — DOI: 10.1016/j.jbi.2020.103496
8. Mikulik I.I. Metodika dlya resheniya zadachi vy'bora priznakov v regressionnoj modeli Koksa [Methodology of solving the feature selection problem for the Cox regression model]. / I.I. Mikulik // Bulletin of Astrakhan State Technical University. Series: Management, Computing Technology and Informatics. — 2024. — № 3. — P. 85–94. — DOI: 10.24143/2072-9502-2024-3-85-94 [in Russian]
9. Davidson-Pilon C. lifelines: survival analysis in Python. / C. Davidson-Pilon // Journal of Open Source Software. — 2019. — № 40. — DOI: 10.21105/joss.01317
10. Faghiri F. Cox proportional hazards model with Bayesian neural network for survival prediction. / F. Faghiri, A. Kohansal // Scientific Reports. — 2025. — № 1. — DOI: 10.1038/s41598-025-16993-4