

**ПРИМЕНЕНИЕ АНСАМБЛЕВЫХ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ
ДЛЯ ПРОГНОЗИРОВАНИЯ ТЕХНИЧЕСКОГО СОСТОЯНИЯ ОБЪЕКТА**

© 2021 Ю.Е. Кувайскова

Ульяновский государственный технический университет, Ульяновск, Россия

Статья поступила в редакцию 10.12.2020

Для обеспечения надёжного функционирования технического объекта необходимо прогнозирование его состояния на предстоящий интервал времени. Пусть техническое состояние объекта характеризуется в определённый момент времени набором параметров, установленных технической документацией на объект. Предполагается, что при определённых значениях этих параметров объект может находиться в исправном или неисправном состоянии. Требуется по значениям этих параметров оценить, в каком состоянии будет находиться объект в предстоящий интервал времени. Для решения этой задачи могут быть применены методы машинного обучения с учителем. Однако для получения хороших результатов прогнозирования состояния объекта необходимо правильно выбрать модель обучения. Одним из недостатков моделей машинного обучения является высокое смещение и слишком большой разброс. В данной работе для уменьшения разброса модели предлагается применение ансамблевых методов машинного обучения, а именно процедуры бэггинга. Основная идея ансамбля методов состоит в том, что при правильном сочетании слабых моделей можно получить более точные и устойчивые модели. Целью бэггинга является создание ансамблевой модели, которая является более надёжной, чем отдельные модели, её составляющие. Одним из больших преимуществ бэггинга является его параллелизм, поскольку различные модели ансамбля обучаются независимо друг от друга. Эффективность предлагаемого подхода показана на примере прогнозирования технического состояния объекта по восьми параметрам его функционирования. Для оценки эффективности применения ансамблевых методов машинного обучения для прогнозирования технического состояния объекта используются критерии качества бинарной классификации: точность, полнота и F -мера. Показано, что применение ансамблевых методов машинного обучения позволяет повысить точность прогнозирования состояния технического объекта на 4%–9% по сравнению с базовыми методами машинного обучения. Данный подход может быть использован специалистами для прогнозирования технического состояния объектов во многих технических приложениях, в частности, в авиации.

Ключевые слова: ансамблевые методы, бэггинг, машинное обучение, прогнозирование, технический объект.

DOI: 10.37313/1990-5378-2021-23-1-111-114

ВВЕДЕНИЕ

Прогнозирование технического состояния объекта является неотъемлемой частью обеспечения его функциональной надёжности и направлено на предсказание его работоспособности, а также наступления предельного состояния [1-2].

Пусть техническое состояние объекта в определённый момент времени характеризуется набором параметров его функционирования X_1, X_2, \dots, X_p . Предполагается, что при определённых значениях этих параметров объект может находиться в исправном ($Y = 1$) или неисправном ($Y = 0$) состоянии. Известные результаты измерений параметров функционирования объекта X_1, X_2, \dots, X_p и соответствующих им состояний объекта Y образуют обучающую выборку. Задача сводится к поиску зависимости (модели) между параметрами объекта X_1, X_2, \dots, X_p и состояниями Y , способной с достаточ-

Кувайскова Юлия Евгеньевна, кандидат технических наук, доцент кафедры «Прикладная математика и информатика». E-mail: u.kuvaiskova@mail.ru

но высокой точностью дать прогноз технического состояния объекта (исправен или неисправен) в предстоящий интервал времени.

Для решения этой задачи могут быть использованы различные модели временных рядов [3-7], модели изображений [8-9], нечёткие модели [10-12], а также методы машинного обучения [13-20]. Однако при использовании методов машинного обучения нельзя заранее определить, какой из них обеспечит корректное решение задачи. Поэтому при решении конкретных задач проводят апробацию множества методов машинного обучения, что иногда бывает трудоёмко. В данной работе для решения поставленной задачи предлагается применение ансамблевых методов машинного обучения [15-20], а именно процедуры бэггинга [18-20], что позволит улучшить точность прогнозирования состояния объекта.

Целью работы является исследование эффективности применения ансамблевых методов машинного обучения для прогнозирования технического состояния объекта.

АНСАМБЛЕВЫЕ МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

Для получения хороших результатов прогнозирования технического состояния объекта необходимо правильно выбрать модель обучения. Выбор модели может зависеть от многих факторов: количества исходных данных, значимости параметров функционирования объекта, наличия корреляционной зависимости между параметрами объекта и т.д. Базовые модели машинного обучения работают по отдельности не всегда хорошо, в связи с тем, что они либо имеют высокое смещение, либо слишком большой разброс, чтобы быть устойчивыми. Поэтому при выборе модели прогнозирования желательно, чтобы она имела слабое смещение и небольшой разброс. Идея ансамблевых методов машинного обучения состоит в том, чтобы попытаться уменьшить смещение и (или) разброс таких слабых моделей, объединяя несколько из них вместе, для получения лучших результатов прогнозирования, чем могли бы получить от каждой модели по отдельности.

В ансамблевых методах обучения базовые модели используются в качестве строительных блоков для проектирования более сложных моделей путём объединения нескольких из них. Для реализации ансамблевых методов нужно отобрать базовые модели для агрегирования. Одним из ансамблевых методов является бэггинг [18–20].

Бэггинг – это один из первых и самых простых видов ансамблей. В методе бэггинга используется единственный базовый алгоритм обучения, то есть имеются однородные слабые модели, которые обучаются параллельно и независимо друг от друга, а затем объединяются, следуя некоторому детерминированному процессу усреднения. Получаемая таким образом модель ансамбля называется «однородной». В основном бэггинг направлен на получение ансамблевой модели с меньшим разбросом, чем её компоненты. Бэггинг основан на статистическом методе бутстрэпа, который позволяет оценивать многие статистики сложных распределений [19].

Бутстрэп – статистический метод, заключающийся в генерации выборок размера B из исходного множества наблюдений размера N путём случайного выбора элементов с повторениями. При некоторых допущениях эти выборки имеют довольно хорошие статистические свойства: в первом приближении их можно рассматривать как взятые непосредственно из истинного базового (и часто неизвестного) распределения данных, так и независимо друг от друга. Таким образом, их можно считать репрезентативными и независимыми выборками. При этом размер N

исходной выборки данных должен быть достаточно большим, чтобы охватить большую часть сложности базового распределения, а также достаточно большим по сравнению с размером бутстрэп выборок B , чтобы выборки не слишком сильно коррелировали.

Идея бэггинга в таком случае проста: необходимо подобрать несколько независимых моделей и «усреднить» их прогнозы, чтобы получить модель с меньшим разбросом.

Пусть сформировано L бутстрэп выборок размера B . По каждой выборке обучено L базовых моделей и получены вероятности для каждого класса Y , предсказываемые всеми моделями:

$$\hat{P}_1(Y), \hat{P}_2(Y), \dots, \hat{P}_L(Y), \quad (1)$$

где $\hat{P}_l(Y), l = \overline{1, L}$ – прогнозируемая вероятность принадлежности объекта к классу Y .

Затем полученные модели объединяются с помощью процедуры усреднения, чтобы получить модель ансамбля с меньшим разбросом, чем её отдельные компоненты:

$$\hat{P}_{bag}(Y) = \frac{1}{L} \sum_{l=1}^L \hat{P}_l(Y). \quad (2)$$

Так как бутстрэп выборки являются примерно независимыми и одинаково распределёнными, то и «усреднение» результатов базовых моделей не изменяет ожидаемый ответ, но уменьшает его разброс (так же, как усреднение независимых одинаково распределённых случайных величин сохраняет ожидаемое значение, но уменьшает разброс).

Одним из больших преимуществ бэггинга является его параллелизм. Поскольку различные модели обучаются независимо друг от друга, при необходимости могут использоваться методы интенсивного распараллеливания.

КРИТЕРИИ КАЧЕСТВА ПРОГНОЗИРОВАНИЯ

Для оценки точности прогнозирования технического состояния объекта с помощью ансамблевых моделей машинного обучения можно использовать критерии качества бинарной классификации [14].

При несбалансированном распределении наблюдений за объектом на классы состояний (исправные и неисправные) используются такие метрики, как точность, полнота и F -мера.

Точность показывает, какая доля наблюдений за объектом, отнесённых моделью к исправному состоянию, действительно принадлежат исправному состоянию:

$$P = \frac{TP}{TP + FP}, \quad (3)$$

где TP – количество правильно классифицированных исправных состояний; FP – количество

неправильно классифицированных исправных состояний.

Полнота – это доля найденных моделью наблюдений исправных состояний объекта относительно всех наблюдений, принадлежащих исправному состоянию, в обучающей выборке:

$$R = \frac{TP}{TP + FN}, \quad (4)$$

где TP – количество правильно классифицированных исправных состояний; FN – количество неправильно классифицированных неисправных состояний объекта.

F-мера – это гармоническое среднее точности и полноты:

$$F = \frac{2PR}{P + R}. \quad (5)$$

Чем ближе значение F к единице, тем лучше результаты классификации.

ЧИСЛЕННОЕ ИССЛЕДОВАНИЕ

В качестве объекта исследования рассмотрим объект, техническое состояние которого определяется восемью параметрами его функционирования.

Для оценки точности прогнозирования состояния объекта разделим исходную выборку на две части: обучающую (90% от объема исходных наблюдений) и контрольную (оставшиеся 10% наблюдений).

По обучающей выборке построим модели базовых методов машинного обучения (логистическая регрессия, байесовский классификатор, дискриминантный анализ, метод опорных векторов, деревья решений, нейронные сети), а также с помощью бэггинг-подхода.

Рассчитаем значение F-меры, полученные результаты исследования представим в таблице 1.

Из таблицы следует, что применение процедуры бэггинга позволяет повысить точность прогнозирования в среднем на 4%–9% по сравнению с базовыми моделями. При этом лучшие результаты дают бэггинг деревьев решений и бэггинг нейронных сетей.

ЗАКЛЮЧЕНИЕ

Для прогнозирования технического состояния объекта предложено использование ансамблевых методов машинного обучения, достоинствами которых является уменьшение разброса модели и получение более надежных результатов прогнозирования.

Эффективность бэггинг-подхода продемонстрирована на примере прогнозирования состояния технического объекта, которое описывается восемью параметрами его функционирования. Оказалось, что применение бэггинга позволяет повысить точность прогнозирования на 4%–9% по сравнению с базовыми методами машинного обучения. Одним из больших преимуществ бэггинга является его параллелизм, поскольку различные модели ансамбля обучаются независимо друг от друга.

Описанный подход может быть использован специалистами для прогнозирования технического состояния объектов во многих технических приложениях, в частности, в авиации.

СПИСОК ЛИТЕРАТУРЫ

1. Биргер И.А. Техническая диагностика. М.: Машиностроение, 1978. 240 с.
2. Клячкин В.Н., Жуков Д.А. Алгоритм диагностики функционирования технического объекта с использованием агрегированных классификаторов // Автоматизация процессов управления. 2019. № 2 (56). С. 37-43.
3. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. М.: Мир, 1974. 242 с.
4. Stock, J.H., Watson M.W. Vector Autoregressions // Journal of Economic Perspectives. 2001. V. 15. P.101-115.
5. Валеев С.Г. Регрессионное моделирование при обработке данных. Казань: ФЭН, 2001. 296 с.
6. Родионова Т.Е. Применение адаптивного регрессионного моделирования для описания функционирования технического объекта // Известия Самарского научного центра Российской академии наук. 2014. Т. 16. № 6-2. С. 572-575.
7. Кадырова Г.Р. Оценка и прогнозирование состояния технического объекта по регрессионным моделям // Автоматизация процессов управления. 2015. № 4 (42). С. 90-95.

Таблица 1. Критерии качества прогнозирования

Метод машинного обучения	F-мера	
	Базовая модель	Бэггинг
Логистическая регрессия	0,8333	0,8710
Байесовский классификатор	0,7719	0,8525
Дискриминантный анализ	0,8136	0,9062
Метод опорных векторов	0,7931	0,8889
Деревья решений	0,9362	0,9855
Нейронные сети	0,9231	0,9855

8. Васильев К.К., Крашенинников В.Р. Статистический анализ последовательностей изображений. М.: Радиотехника, 2017. 248 с.
9. Krashennnikov V.R., Kuvayskova Y.E. Modelling and Forecasting of Quasi-Periodic Processes in Technical Object Based on Cylindrical Image Models // CEUR Workshop Proceedings. DS-ITNT 2019 - Proceedings of the Data Science Session at the 5th International Conference on Information Technology and Nanotechnology. 2019. P. 387-393.
10. Mamdani E.H. Application of fuzzy algorithms for control of simple dynamic plants // Proceedings of the Institution of Electrical Engineers. 1974. V. 121, No. 12. P. 1585-1588.
11. Kuvayskova Y.E. The Prediction Algorithm of the Technical State of an Object by Means of Fuzzy Logic Inference Models // Procedia Engineering. «3rd International Conference «Information Technology and Nanotechnology», ITNT 2017». 2017. P. 767-772.
12. Кувайскова Ю.Е. Использование нечёткой логики для диагностики технического состояния объекта // Известия Самарского научного центра Российской академии наук. 2018. Т. 20. № 4-3 (84). С. 487-490.
13. Witten I.H., Frank E. Data mining: practical machine learning tools and techniques. 2nd ed. San Francisco: Morgan Kaufmann Publishers, 2005. 525 p.
14. Клячкин В.Н., Жуков Д.А. Прогнозирование состояния технического объекта с применением методов машинного обучения // Программные продукты и системы. 2019. № 2. С. 244-250.
15. Explaining the success of AdaBoost and random forests as interpolating classifiers / A.J. Wyner, M. Olson, J. Bleich, D. Mease // The Journal of Machine Learning Research. 2017. No. 18. P. 1-33.
16. Жуков Д.А., Клячкин В.Н. Использование агрегированных классификаторов при машинном обучении в задачах технической диагностики // Информационные технологии в моделировании и управлении: подходы, методы, решения. Материалы II Всероссийской научной конференции с международным участием. 2019. С. 76-81.
17. Selection of aggregated classifiers for the prediction of the state of technical / D.A. Zhukov, V.N. Klyachkin, V.R. Krashennnikov, Yu.E. Kuvayskova // CEUR Workshop Proceedings. DS-ITNT 2019 - Proceedings of the Data Science Session at the 5th International Conference on Information Technology and Nanotechnology. 2019. С. 361-367.
18. Breiman L. Bagging Predictors // Machine Learning. 1996. V. 24 (2). P. 123-140.
19. Открытый курс машинного обучения. Композиции: бэггинг, случайный лес. URL: <https://habr.com/ru/company/ods/blog/324402/> (дата обращения: 17.11.2020)
20. Евсеева С.А. Исследование эффективности процедуры бэггинга в коллективном принятии решения // Решетневские чтения. Материалы XXIII Международной научно-практической конференции, посвященной памяти генерального конструктора ракетно-космических систем академика М.Ф. Решетнева. Под общей редакцией Ю.Ю. Логинова. 2019. С. 159-160.

APPLICATION OF ENSEMBLE MACHINE LEARNING METHODS FOR PREDICTING THE TECHNICAL STATE OF AN OBJECT

© 2021 Yu.E. Kuvayskova

Ulyanovsk State Technical University, Ulyanovsk, Russia

To ensure the reliable functioning of a technical object, it is necessary to predict its state for the upcoming time interval. Let the technical state of the object be characterized at a certain point in time by a set of parameters established by the technical documentation for the object. It is assumed that for certain values of these parameters, the object may be in a good or faulty state. It is required by the values of these parameters to estimate the state of the object in the upcoming time interval. Supervised machine learning methods can be applied to solve this problem. However, to obtain good results in predicting the state of an object, it is necessary to choose the correct training model. One of the disadvantages of machine learning models is high bias and too much scatter. In this paper, to reduce the scatter of the model, it is proposed to use ensemble machine learning methods, namely, the bagging procedure. The main idea of the ensemble of methods is that with the right combination of weak models, more accurate and robust models can be obtained. The purpose of bagging is to create an ensemble model that is more reliable than the individual models that make up it. One of the big advantages of bagging is its concurrency, since different ensemble models are trained independently of each other. The effectiveness of the proposed approach is shown by the example of predicting the technical state of an object by eight parameters of its functioning. To assess the effectiveness of the application of ensemble machine learning methods for predicting the technical state of an object, the quality criteria of binary classification are used: accuracy, completeness, and F-measure. It is shown that the use of ensemble machine learning methods can improve the accuracy of predicting the state of a technical object by 4%–9% in comparison with basic machine learning methods. This approach can be used by specialists to predict the technical condition of objects in many technical applications, in particular, in aviation.

Keywords: ensemble methods, bagging, machine learning, predicting, technical object.

DOI: 10.37313/1990-5378-2021-23-1-111-114

Yuliya Kuvayskova, Candidate of Technics, Associate Professor at the Applied Mathematics and Informatics Department. E-mail: u.kuvayskova@mail.ru