

Анализ стабильности функционирования объекта с применением машинного обучения

В.Н. Клячкин¹, Д.А. Жуков¹, Е.А. Зенцова¹

¹Ульяновский государственный технический университет, Северный Венец 32, Ульяновск, Россия, 432027

Аннотация. Стабильность функционирования технического объекта оценивается методами статистического управления процессами. Однако не всегда такой подход обеспечивает своевременное обнаружение нарушения. Предложено использование методов машинного обучения для бинарной классификации состояния объекта (стабилен или нестабилен). Разработана программа для проведения расчетов в системе Matlab, обеспечивающая анализ влияния метода обучения, критериев качества классификации, способа формирования контрольной выборки, а также способов отбора значимых показателей, на точность прогнозирования стабильности функционирования объекта. В качестве примеров рассмотрены стабильность работы системы управления водоочисткой, стабильность вибраций гидроагрегата, технологический процесс механической обработки.

1. Введение

Стабильность функционирования технического объекта оценивается методами статистического управления процессами. Однако не всегда такой подход обеспечивает своевременное обнаружение нарушения [1-2]. Предлагается использование методов машинного обучения для бинарной классификации состояния объекта (стабилен или нестабилен). Вместе с тем, накопленные результаты статистического контроля могут быть использованы в качестве исходных данных для прогнозирования состояния объекта с применением методов машинного обучения: для каждой выборки известны значения контролируемых показателей и состояние функционирования объекта.

Качество бинарной классификации существенно зависит от ряда факторов. Во-первых, это выбранный метод машинного обучения: могут быть использованы как базовые (наивный байесовский классификатор, нейронная сеть и другие), так и композиционные (бэггинг или бустинг различных видов), а также агрегированные методы [3-6]. Во-вторых, на качество диагностики влияет выбранный критерий: наиболее распространены доля ошибок на контрольной выборке, F-критерий и площадь AUC под ROC-кривой [7-8]. Для технических объектов характерны выборки с небольшим объемом неисправных состояний. В этом случае более эффективен F-критерий. На результаты обучения оказывают влияние также методы отбора значимых показателей функционирования объекта (использование малозначимых показателей может привести к некорректным результатам), а также способ формирования и объем контрольной выборки [9].

Цель исследования – разработка программы для оценки стабильности функционирования технического объекта, обеспечивающей анализ влияния метода машинного обучения, способа формирования контрольной выборки, а также отбора значимых показателей, на точность прогнозирования стабильности, а также проведение численного анализа стабильности функционирования на примере реальных объектов.

2. Подготовка исходных данных

Статистическое управление процессом состоит в выявлении неслучайных нарушений, связанных с так называемыми особыми причинами [10-12]. Например, для процесса механической обработки такими причинами могут быть износ резца или ослабление его крепления, изменения в составе смазочно-охлаждающей жидкости и т.п. Основное преимущество такого подхода в том, что нарушение выявляется до того, как контролируемый показатель вышел за допустимые границы, при этом статистические методы уже показывают нарушение стабильности процесса.

Для многопараметрического процесса вначале исследуются корреляционные связи между показателями функционирования объекта, затем выбираются средства контроля и определяются их параметры. Для независимых показателей используют карты Шухарта (контролируется средний уровень процесса и его рассеяние). Стабильность коррелированных показателей оценивают с применением алгоритмов Хотеллинга и обобщенной дисперсии [13-15]. Причины нарушений при многомерном контроле процесса могут быть найдены с использованием частного критерия Хотеллинга.

Предположим, что проведено исследование N выборок по d показателям функционирования, при этом в k выборках контрольные карты выявили нарушение стабильности процесса. Таким образом формируется множество N прецедентов $(x^{(i)}, y^{(i)})$, $i = 1 \dots N$: объекты с заданными d показателями функционирования $x = (x_1, x_2, \dots, x_d)$ и соответствующими состояниями y , принимающими одно из двух значений $(0, 1)$; $y = 0$ – соответствует нестабильному состоянию (таких прецедентов k), $y = 1$ стабильному ($N - k$ прецедентов). На основе этих данных надо восстановить зависимость между показателями функционирования и состоянием объекта.

3. Диагностика состояния объекта

Практическая реализация методов машинного обучения возможна на базе библиотеки инструментов Statistics and Machine Learning Toolbox в пакете Matlab. С учетом целей исследования была разработана программа, обеспечивающая:

- использование различных базовых и композиционных методов, а также построение агрегированных классификаторов трех типов: с агрегированием по среднему значению, по медиане или голосованием [16],
- применение различных критериев качества классификации: доли ошибок на контрольной выборке, F-критерия и площади AUC под ROC-кривой,
- отбор значимых показателей функционирования объекта путем построения регрессионной модели для зависимости состояния объекта y от показателей функционирования x_j ($j = 1 \dots d$) и проверки значимости показателей по критерию Стьюдента,
- варьирование способа формирования контрольной выборки (случайный отбор или определенная часть исходных данных) и ее объем [9].

Для проведения диагностики стабильности состояния технического объекта вводится выборка исходных данных, проводится машинное обучение по всем встроенным в Matlab базовым и композиционным методам; при этом качество классификации оценивается с применением кросс-валидации по F-критерию. Отбирают значимые показатели функционирования и проверяют, улучшает ли эта процедура качество модели. Варьируя объем контрольной выборки от 5 до 25% от исходной, выбирают объем, обеспечивающий наилучшее значение критерия. Для этого варианта строятся агрегированные классификаторы и выбирается классификатор, обеспечивающий максимум F-критерия.

4. Численное исследование

Статистический контроль стабильности процесса механической обработки (шлифования четырех цилиндрических поверхностей) оси проведен по 400 выборкам. Все четыре контролируемых показателя (диаметры ступеней оси) оказались коррелированы между собой. Для контроля среднего уровня процесса использовалась карта Хотеллинга, для контроля рассеяния – карта обобщенной дисперсии. На рисунке 1 показаны результаты статистического контроля для первых 51 выборок. Видно, что нарушения процесса (выход за контрольную границу) на карте Хотеллинга имеет место в трех выборках, карте обобщенной дисперсии – в двух.

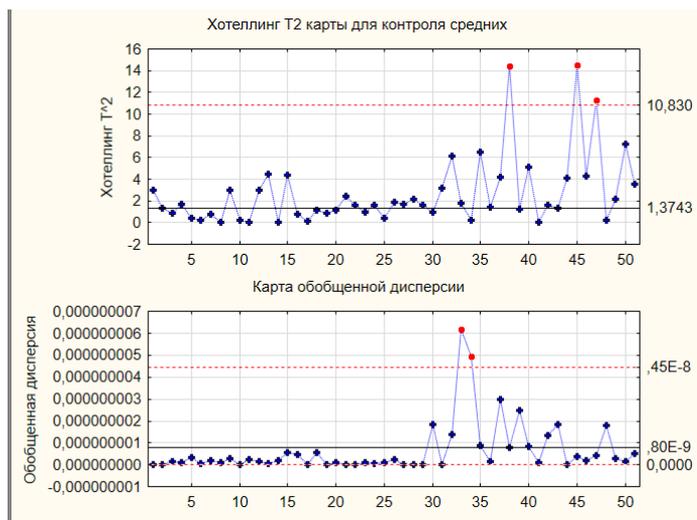


Рисунок 1. Карты Хотеллинга и обобщенной дисперсии.

Таблица исходных данных была подготовлена по всем 400 наблюдениям. При проведении исследования с помощью разработанной программы было установлено, что все четыре показателя значимы, максимальное значение F-критерия 0,874 было получено для контрольной выборки объемом 15% от исходной выборки при применении агрегированного классификатора по среднему значению, включающему метод опорных векторов и бэггинг деревьев решений. При этом значение F-критерия увеличилось на 8%. В качестве стандартного подхода рассмотрен наилучший классификатор по отдельным методам (здесь – бэггинг деревьев решений), все показатели значимы.

Во втором испытании оценивалась стабильность вибраций гидроагрегата по результатам показаний 10 датчиков в 5000 наблюдений [17]. Множество их 10 показателей разделилось на четыре подмножества: два показателя оказались некоррелированными с другими, в третью группу вошли три коррелированных показателя, в четвертую – пять. Первые два независимых показателя контролировались с помощью карт Шухарта для средних значений и стандартных отклонений. Для третьей и четвертой групп строились карты Хотеллинга и обобщенной дисперсии. По результатам статистического контроля формировалась выборка исходных данных, по которой и проводилось машинное обучение. Значимыми оказались семь из десяти показателей. Наилучшим оказался агрегированный классификатор по медиане, включающий градиентный бустинг и логистическую регрессию, объем контрольной выборки 20%. Увеличение F-критерия (до значения 0,904) составило 18%.

При анализе стабильности функционирования системы водоочистки по восьми показателям качества питьевой воды использовались результаты 1557 наблюдений (исправное состояние в 1204 случаях). Значимыми оказались шесть показателей. Максимальное значение F-критерия имело место при агрегировании нейронной сети и бэггинга деревьев решений; в этом опыте увеличение значения критерия по сравнению с нейронной сетью (лучший отдельный классификатор) незначительно: с 0,879 до 0,881. При этом объем контрольной выборки составил 10%.

5. Заключение

Проведенное исследование показало, что точность диагностирования стабильности функционирования технического объекта с применением бинарной классификации может быть повышена (в рассмотренных примерах до 18%) за счет агрегирования методов, отбора значимых показателей и изменения объема контрольной выборки. Разработанная программа обеспечивает автоматический поиск наилучшего варианта. При этом в качестве исходных данных могут использоваться результаты статистического контроля процесса функционирования рассматриваемого объекта.

6. Литература

- [1] Клячкин, В.Н. Статистические методы оценки стабильности функционирования технических систем / В.Н. Клячкин, И.Н. Карпунина // Надежность и качество сложных систем. – 2018. – № 2. – С. 36-42
- [2] Клячкин, В.Н. Система статистического контроля многопараметрического технологического процесса // Приборы и системы. Управление, контроль, диагностика. – 2008. – № 10 – С. 30-33.
- [3] Witten, I.H. Data mining: practical machine learning tools and techniques / I.H. Witten, E. Frank. – San Francisco: Morgan Kaufmann Publishers, 2005. – 525 p.
- [4] Мерков, А.Б. Распознавание образов. Введение в методы статистического обучения / А.Б. Мерков. – М.: Едиториал УРСС, 2011. – 256 с.
- [5] Воронина, В.В. Теория и практика машинного обучения : учеб. пособие / В.В. Воронина, А.В. Михеев, Н.Г. Ярушкина, К.В. Святков. – Ульяновск : УлГТУ, 2017. – 290 с.
- [6] Воронцов, К.В. Машинное обучение. Композиция классификаторов [Электронный ресурс]. – Режим доступа: <https://yadi.sk/i/FItIu6V0beBmF> (01.11.2018).
- [7] Соколов, Е.А. ФКН ВШЭ. Лекция 4. Линейная классификация [Электронный ресурс]. – Режим доступа: <https://github.com/esokolov/ml-course-hse/blob/master/2018-fall/lecture-notes/lecture04-linclass.pdf> (01.11.2018).
- [8] Дьяконов, А.М. AUC ROC (площадь под кривой ошибок) [Электронный ресурс]. – Режим доступа: <https://dyakonov.org/2017/07/28/auc-roc-площадь-под-кривой-ошибок/#more-5362> (01.11.2018).
- [9] Жуков, Д.А. Влияние объема контрольной выборки на качество диагностики состояния технического объекта / Д.А. Жуков, В.Н. Клячкин // Автоматизация процессов управления. – 2018. – № 2. – С. 90-95.
- [10] Уиллер, Д. Статистическое управление процессами Оптимизация бизнеса с использованием контрольных карт Шухарта / Д. Уиллер, Д. Чамберс. – М.: Альпина Бизнес Букс, 2009. – 409 с.
- [11] Montgomery, D.C. Introduction to statistical quality control. – New York: John Wiley and Sons, 2009. – 754 p.
- [12] Клячкин, В.Н. Статистические методы в управлении качеством: компьютерные технологии. – М.: Финансы и статистика, ИНФРА-М, 2009. – 304 с.
- [13] Клячкин, В.Н. Диагностика многопараметрического технологического процесса с использованием контрольных карт на главных компонентах / В.Н. Клячкин, Е.И. Константинова // Приборы и системы. Управление, контроль, диагностика. – 2007. – № 3. – С. 59-61.
- [14] Святова, Т.И. Многомерный статистический контроль технологического рассеяния процесса / Т.И. Святова, В.Н. Клячкин // Радиотехника. – 2014. – №11. – С. 123-126.
- [15] Зенцова, Е.А. Адаптивный многомерный статистический контроль технологического процесса / Е.А. Зенцова, В.Н. Клячкин // Радиотехника. – 2017. – № 6. – С. 45-48.
- [16] Klyachkin, V.N. The use of aggregate classifiers in technical diagnostics, based on machine learning / V.N. Klyachkin, Yu.E. Kuvayskova, D.A. Zhukov // CEUR Workshop Proceeding. – 2017. – Vol. 1903. – P. 32-35.

[17] Клячкин, В.Н. Система статистического анализа и контроля стабильности вибраций гидроагрегата / В.Н. Клячкин, Ю.Е. Кувайскова, А.В. Иванова // Программные продукты и системы. – 2018. – № 3. – С. 600-625.

Благодарности

Исследование выполнено при финансовой поддержке РФФИ и Правительства Ульяновской области, проект 18-48-730001.

Analysis of stable functioning of objects using machine learning

V.N. Klyachkin¹, D.A. Zhukov¹, E.A. Zentsova¹

¹Ulyanovsk State Technical University, Severny Venec street 32, Ulyanovsk, Russia, 432027

Abstract. Stable functioning of technical objects is evaluated by methods of statistical process control. However such approach not always provides timely detection of malfunction. It has been suggested that methods of machine learning should be used for binary classification of the object state (stable or unstable). A program has been developed for calculation in the Matlab environment which allows for analysis of impact of the learning method, classification quality criteria, method of validation set as well as methods of selection of significant indicators on the object's stable functioning forecast precision. Stable operation of the water treatment management system, stable vibration of the hydraulic unit, machining operation process are taken as examples.