## Обнаружение разладки технологического сигнала на основе классификации

**Гаврин Сергей Владимирович,** магистрант первого года обучения на магистерской программе "Интеллектуальный анализ больших данных" ТГУ, Томск, тел: 89832302507, [gavrin\_s@mail.ru](mailto:gavrin_s@mail.ru).

**Мурзагулов Дамир Альбертович**, ведущий инженер автоматизированных систем управления технологическими процессами группы компаний Imperial Energy Ltd, аспирант кафедры теоретических основ информатики ТГУ, Томск, тел: 89138177056, [damir.murzagulov@imperialenergy.ru](mailto:damir.murzagulov@imperialenergy.ru).

**Замятин Александр Владимирович**, докт. техн. наук, доцент, заведующий кафедрой теоретических основ информатики ТГУ, Томск, тел: 89069576374, [zamyatin@mail.tsu.ru](mailto:zamyatin@mail.tsu.ru).

# Введение

Успехи последних лет в области искусственного интеллекта и машинного обучения [1] привели к попыткам создания промышленных систем предиктивной аналитики, предназначенных для прогнозирования отказов оборудования и аварийных ситуаций на производстве. В основе подобных систем лежит обработка и анализ исторических данных, поступающих с автоматизированных систем управления технологическими процессами (АСУ ТП) в виде временных рядов [2]. Многие проблемы анализа и управления сложными технологическими процессами сводятся к обнаружению резких изменений (разладки) характеристик анализируемого временного ряда [3]. Разладка представляет собой нарушение закономерности в динамике или структуре временного ряда и может служить предвестником критических явлений и аварийных ситуаций [4][5]. Автоматическая идентификация разладки затрудняется в связи со сложной структурой технологических сигналов, большим количеством компонент и локальных особенностей сигнала.

На рис. 1 изображён тренд изменения уровня нефти в сепараторе за сутки. С 8 до 14 часов можно наблюдать два момента разладки сигнала, выделенные красными окружностями. Стоит заметить, что указать точное время разладки для технологического сигнала довольно затруднительно в условиях быстрой динамики изменения. Прикладное значение имеет определение промежутка времени, когда происходит нарушение закономерности.

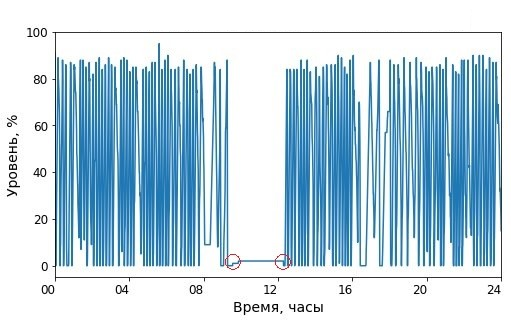


Рис. 1. График уровня перепада давления.

Существует две основные группы методов обнаружения момента разладки, зарекомендовывавших себя для различных видов временных рядов: статистические [6][7] и методы машинного обучения [8]. Несмотря на широкую область применения статистических методов в аналогичных областях анализа временных рядов (биомедицина, финансы и др.), возникает ряд недостатков при применении этих методов для временных рядов техногенного происхождения [12]. К их числу относится, сложный математический аппарат, необходимость адаптации моделей для каждого набора данных, ограничение по стационарности распределения. Данные ограничения на некотором уровне могут быть разрешимы с помощью правильно подобранных моделей машинного обучения. При разработке этих алгоритмов важно учитывать наличие или отсутствие априорной информации о протекании процесса. В зависимости от этого выбираются техники построения алгоритмов: контролируемая или неконтролируемая классификация [1].

Технологические процессы протекают как в установившихся, так и в переходных режимах, которые находят отражение в технологических сигналах в виде определённых паттернов (шаблонов) поведения. Такие паттерны могут иметь схожую структуру, амплитуду, частоту и др. характеристики. Таким образом, сложный технологический сигнал может быть представлен набором последовательных паттернов, соответствующих режимам технологического процесса. В терминах машинного обучения паттерн сигнала можно рассматривать как класс объекта, а модели, определяющие к какому классу относится фрагмент временного ряда, классификаторами.

В работе предлагается разработать вычислительно эффективный алгоритм обнаружения разладки в технологических сигналах при использовании контролируемой классификации и модельных классов сигналов.

## Постановка задачи

Пусть имеется технологический сигнал , который имеет структуру:

,

где, и представляет собой разные классы сигналов, представленные в виде временного ряда с равноотстоящими отсчётами , где - значение ряда в момент времени ; При этом говорят, что соответствует состоянию до разладки (штатному), а -состоянию с разладкой (аномальному); момент принято называть моментом появления разладки. Задача «о разладке» состоит в следующем: пока наблюдения согласуются с нормальным состоянием, требуется продолжать наблюдения. Если состояние изменяется на аномальное, требуется обнаружить изменение как можно скорее, избегая ложных тревог.

В прикладных задачах обнаружения разладки достаточно указать минимально возможный промежуток времени, в который появляется разладка. Поэтому запишем постановку задачи следующим образом. Для анализируемого сигнала определить временной диапазон (), далее допустимый диапазон разладки, для которого выполняются условия:

|  |  |
| --- | --- |
|  | (1) |

где - неизвестный момент разладки, - допустимый момент тревоги.

Цель работы: разработать алгоритм обнаружения разладки технологических сигналов с минимальным возможным моментом тревоги и с использованием априорной информации о классах сигналов.

## Критерии качества

Рассмотрим множество алгоритмов , в рамах которых будем искать решение задачи.

Введём функцию потерь, как отклонения от допустимого диапазона (deviation from range):

|  |  |
| --- | --- |
|  | (2) |

Данная функция равна 0 если момент разладки попадает в найденный диапазон, в других случаях, равна минимальному расстоянию от истинного момента разладки до спрогнозированного диапазона. Для ложных срабатываний алгоритма значение будет значительно превышать .

Для оценки качества алгоритма введём функционал эмпирического риска среднее отклонение оn допустимого диапазона (average deviation from range):

|  |  |
| --- | --- |
|  | (3) |

где – количество экспериментов.

Поиск алгоритма заключается в минимизации функционала ADR, при минимизации момента тревоги .

# Экспериментальные данные

Для обучения и апробации алгоритма требуется обучающая выборка, покрывающая широкий спектр сигналов с различными параметрами. Дополнительно, выборка должна содержать метки принадлежности к определённому классу сигнала.

Для проведения численных экспериментов разработана и сгенерирована обучающая выборка, состоящая из 9 элементарных классов сигналов. В качестве классов используются абстрактные модели временных рядов, часто встречающихся в реальных сигналах. В таблице №1 приведено описание классов.

Таблица №1. Описание модельных сигналов

|  |  |  |  |
| --- | --- | --- | --- |
| № | Название | Описание класса | Аналитическое определение |
| 0 | «norm» | Горизонтальная прямая |  |
| 1 | «cyclic» | Гармонический |  |
| 2 | «gauss» | Гауссова функция (монотонное возрастание) |  |
| 3 | «gauss\_up» | Гауссова функция (монотонное убывание) |  |
| 4 | «finite» | Треугольные импульсы |  |
| 5 | «saw» | Пила |  |
| 6 | «rectangle» | Прямоугольные импульсы |  |
| 7 | «dec\_cyclic» | Переходный процесс с затуханием |  |
| 8 | «inc\_cyclic» | Переходный процесс с расхождением |  |

На практике, в условиях промышленной эксплуатации, сигналы содержат шумовые компоненты. По этой причине сигналы из табл. №1 смоделированы с аддитивным белым шумом с математическим ожиданием равным 0 и стандартным отклонением в диапазоне σ = [0,0; 0,2].

Для исследования влияния шума на результаты работы алгоритма введём соотношение сигнал/шум , где – средняя мощность с полезного сигнала, – средняя мощность зашумленного сигнала [9].

Перед использованием, данные из обучающей выборки были нормированы в диапазон [0;1] по следующей формуле: .

На рис. 2 изображены примеры полученных сигналов с различным значением .

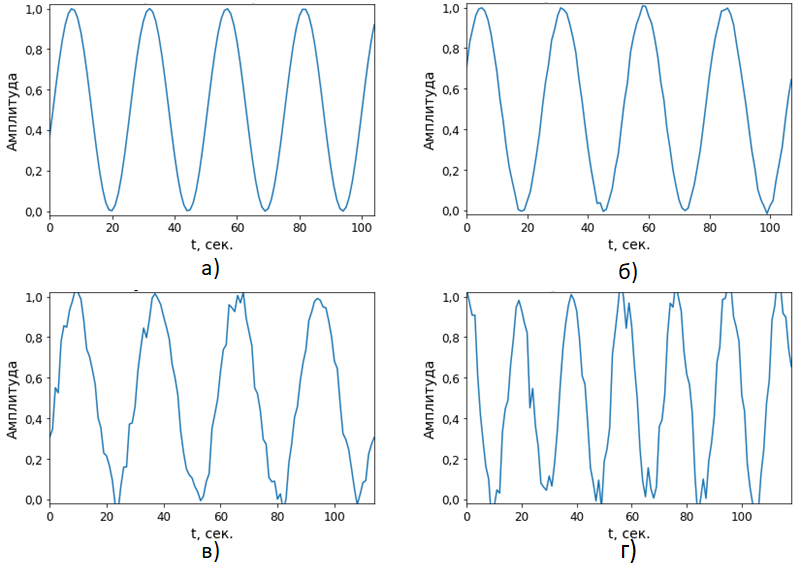


Рис. 2. Примеры модельных сигналов а) Без шума б) SNR = 25.0

в) SNR = 5.0 г) SNR = 3.6

# Предлагаемый алгоритм

Для минимизации функционала и допустимого момента тревоги предлагается использовать каскадную модель классификации с уменьшением входного окна [10]. Такой подход к уточнению момента разладки позволяет плавно, и как правило с минимальной ошибкой, находить область разладки.

Для решения поставленной задачи разработан и апробирован алгоритм обнаружения разладки состоящий из двух основных этапов:

1. На первом этапе определяется класс первой части сигнала .
2. На втором каскадной моделью классификации происходит уточнение границ первой части сигнала.

Дадим словесное описание алгоритма:

1. Задаётся область исследования, равная длине исходного сигнала .
2. Определяется класс первой части сигнала .
3. Запускается цикл по всем классификаторам с уменьшением входного вектора, далее окна :  
   3.1. Производится проход классификатора по области исследования . На выходе получаем ряд спрогнозированных значений классов равный длине .  
   3.2. По полученному ряду, находится область, где изменяется класс сигнала.  
   3.3. Устанавливаются границы в соответствии с найденной областью, равной ширине окна текущего классификатора.
4. Возвращается найденная область , равная минимальному окну классификации, являющаяся решением задачи.

# Описание эксперимента

Для апробации алгоритма проводился эксперимент с применением модельных классов сигналов при различных значениях .

В качестве классификатора в алгоритме использовали свёрточную нейронную сеть FCN [11], обученную на модельных сигналах.

На этапе подготовки данных проводилась «склейка» разных видов сигналов с различными значениями , имитирующая смену режима. Длины всех сигналов выбирались от 90 до 120 отсчетов случайным образом.

Далее на этих данных проводилось по 100 запусков алгоритма для всех видов сигналов и некоторых значений в диапазоне от 25,0 до 1,2 .

# Результаты эксперимента

В ходе проведённого эксперимента был получен ряд результатов.

На рис. 4 изображён график среднего отклонения от допустимого момента разладки на различных видов сигналов при . На рис. 3 изображён график среднего отклонения от допустимого момента разладки в зависимости от некоторых значений .

Как видно из рис. 3, при увеличении шума (уменьшении ), качество алгоритма постепенно ухудшается. Так же стоит заметить, что на таких классах сигналов как gauss, dec\_cyclic и gauss\_up значения значительно больше остальных. То есть обнаружение разладки происходит раньше, чем мы предполагали. Связано это с тем, что некоторые части таких сигналов могут быть похожими на другие, например, norm. Поэтому, можно считать, что в этом случае алгоритм выдаёт ложные срабатывания, хотя, формально, ошибки нет. Для удовлетворительной работы на таких данных требуется дополнительная доработка алгоритма, учитывающая схожесть сигналов между собой.

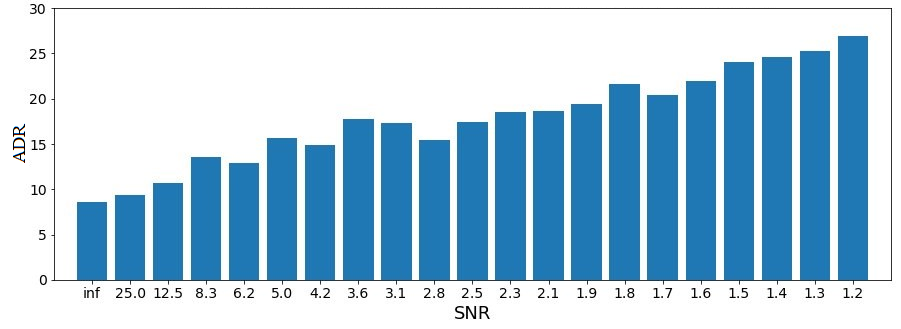


Рисунок 3. Средние отклонения от допустимого момента в зависимости от различных SNR.

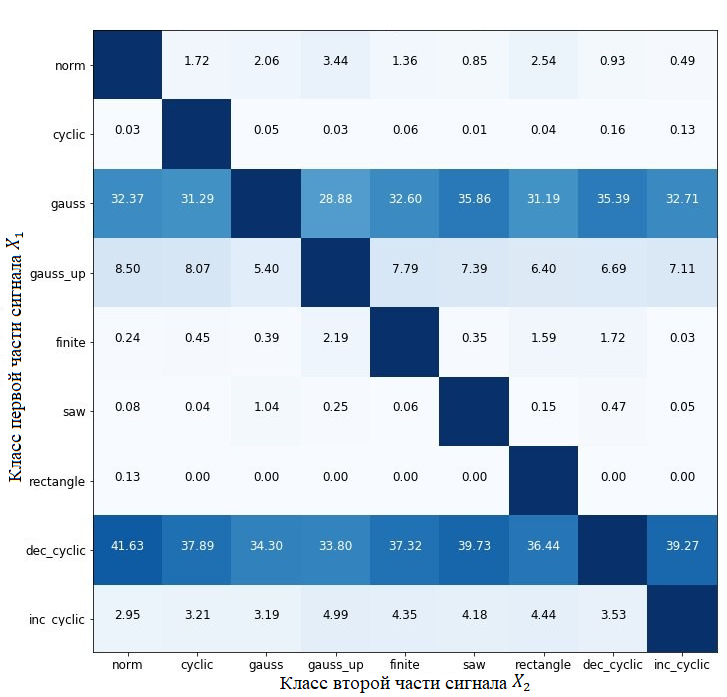


Рисунок 4. Средние отклонения от допустимого момента при SNR = 25.0

# Заключение

В ходе работы произведён обзор методов обнаружения разладки временных рядов. Разработан и реализован оригинальный алгоритм обнаружения разладки для технологических сигналов с априорной информацией о классах сигналов. В основе алгоритма лежит каскадная модель классификация, позволяющая с высокой точностью определять диапазон разладки. Введена метрика для алгоритмов, позволяющая определять временной диапазон разладки. Для обучения и апробации алгоритма была разработана и сгенерирована обучающая выборка модельных сигналов с белым аддитивным шумом. Произведён вычислительный эксперимент на модельных данных и проведена оценка полученных результатов. Предложенный алгоритм показал работоспособность и способность обнаруживать разладки, отражающие переключения режимов процессов или нарушение в технологических процессах. Показано, что в силу своих особенностей, данный алгоритм может быть успешно применён в системах предиктивной аналитики промышленных объектов и других схожих областях.

# Список литературы

[1] Замятин А.В. Введение в интеллектуальный анализ данных: учеб. пособие. – Томск: Издательский Дом государственного университета, 2016. – 120 с.

[2] Замятин А.В., Тренькаев В.Н., Острасть П.М., Телицын Е.А. Высокопроизводительный сервер истории системы диспетчерского управления и сбора данных // ПРОМЫШЛЕННЫЕ АСУ И КОНТРОЛЛЕРЫ. 2017. № 9, С. 20-28. ISSN 1561-1531.

[3] Артемов А. В., Бурнаев Е. В. Исследование процедуры обнаружения разладки временного ряда на основе взвешенного голосования // Материалы Международного молодежного научного форума ЛОМОНОСОВ-2015. — МАКС Пресс (Москва), 2015. — С. 34–35.

[4] Ширяев А. Н. Статистический последовательный анализ. М.: Наука, 1969. Изд-е 2-е: М.: Наука, 1976.

[5] A. G. Tartakovsky, A. S. Polunchenko and G. Sokolov, "Efficient Computer Network Anomaly Detection by Changepoint Detection Methods," in IEEE Journal of Selected Topics in Signal Processing, vol. 7, no. 1, pp. 4-11, Feb. 2013.

[6] Е. В. Бурнаев, “О задаче обнаружения разладки для пуассоновского процесса в обобщенной байесовской постановке”, *Теория вероятн. и ее примен.*, **53**:3 (2008), 534–556; *Theory Probab. Appl.*, **53**:3 (2009), 500–518

[7] Козинов Игорь Александрович, Мальцев Георгий Николаевич Модифицированный алгоритм обнаружения разладки случайного процесса и его применение при обработке многоспектральных данных // Информационно-управляющие системы. 2012. №3 (58).

[8] Artemov A., Burnaev E. Ensembles of detectors for online detection of transient changes // In proc. of Eighth International Conference on Machine Vision. 2015. С. 98751Z–98751Z-5. (Индекс. Scopus)

[9]D. J. Schroeder (1999). Astronomical optics (2nd ed.). Academic Press. p. 433. ISBN 978-0-12-629810-9.

[10] Посохов Иван Александрович, Логунова Оксана Сергеевна, Миков Анатолий Юрьевич, Мацко Игорь Игоревич, Павлов Владимир Викторович Каскадная классификация изображений серного отпечатка поперечного темплета непрерывнолитой заготовки // ЭС и К. 2015. №4 (29).

[11] Z. Wang, W. Yan and T. Oates, "Time series classification from scratch with deep neural networks: A strong baseline," 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, 2017, pp. 1578-1585.

[12] Artemov A., Burnaev E., Lokot A. Nonparametric decomposition of quasi-periodic time series for change-point detection //Eighth International Conference on Machine Vision. — International Society for Optics and Photonics, 2015. — pp. 987520-987520-5.

[13] Боровков А. А. Математическая статистика. Новосибирск: Наука и Изд-во Ин-та математики СО РАН, 1997.

[14] Г. И. Салов, “Задача о разладке для скачкообразного марковского процесса”, *Сиб. журн. индустр. матем.*, **11**:1 (2008), 111–121

[15] А. Н. Колмогоров, Ю. В. Прохоров, А. Н. Ширяев, “Вероятностно-статистические методы обнаружения спонтанно возникающих эффектов”, *Теория вероятностей, теория функций, механика*, Сборник обзорных статей 5. К 50-летию Института, Тр. МИАН СССР, **182**, Наука, М., 1988, 4–23; *Proc. Steklov Inst. Math.*, **182** (1990), 1–21

[16] Г. И. Салов, “Уравнение для апостериорной вероятности наличия «разладки» последовательности зависимых случайных величин и оптимальное по Ширяеву обнаружение момента появления «разладки»”, *Теория вероятн. и ее примен.*, **34**:4 (1989), 799–802; *Theory Probab. Appl.*, **34**:4 (1989), 732–735

[17] *Перов, А. И.* Статистическая теория радиотехнических систем. — М.: Радиотехника, 2003. — 400 с. — ISBN 5-93108-047-3.

[18] A. S. Polunchenko, “Asymptotic near-minimaxity of the randomized Shiryaev–Roberts–Pollak change-point detection procedure in continuous time”, *Теория вероятн. и ее примен.*, **62**:4 (2017), 769–786

[19] Artemov A., Burnaev E., Lokot A. Nonparametric decomposition of quasi-periodic time series for change-point detection //Eighth International Conference on Machine Vision. — International Society for Optics and Photonics, 2015. — pp. 987520-987520-5.

[20] Zamyatin, Alexander. (2015). Data Science and High Energy Physics collaboration enforcement by Higher Education Institutions // DOI: Zenodo. 10.5281/zenodo.3571  
 [21] Kobyz G.V, Zamyatin A.V. Conditional Probability Density Estimation Using Artificial Neural Network // 9th IEEE International Conference on Application of Information and Communication Technologies (AICT), 14-16 Oct 2015, Rostov-on-Don, pp. 441-445. ISBN: 978-1-4673-6855-1

[22] Daniel Stamate, Alexander Zamyatin etc. A Novel Decision Tree Algorithm Based on Parameterised Impurities and Statistical Pruning Approaches // IFIP Advances in Information and Communication Technology, AIAI 2018, 25-27 May 2018, 14th International Conference on Artificial Intelligence Applications and Innovations (в печати)

[23]Daniel Stamate, Wajdi Alghamdi, Daniel Stahl, Alexander Zamyatin, Robin Murray and Marta di Forti. Can Artificial Neural Networks Predict Psychiatric Conditions Associated with Cannabis Use? // AIAI 2018, 25-27 May 2018, 14th International Conference on Artificial Intelligence Applications and Innovations (в печати)

[24] Polunchenko A. S., Sokolov G., Du W. Quickest change-point detection: A bird's eye view //arXiv preprint arXiv:1310.3285. – 2013.

[25]Aminikhanghahi S, Cook DJ. A Survey of Methods for Time Series Change Point Detection. *Knowledge and information systems*. 2017;51(2):339-367