

ЭКСПЛУАТАЦИЯ ОБЪЕКТОВ
АТОМНОЙ ОТРАСЛИ
OPERATION OF FACILITIES
NUCLEAR INDUSTRY

УДК 621.039.564.2:534.2
doi: 10.26583/gns-2022-03-04

**КЛАСТЕРИЗАЦИЯ АКУСТИЧЕСКИХ СОБЫТИЙ В ГЛАВНОМ
ЦИРКУЛЯЦИОННОМ КОНТУРЕ РЕАКТОРНОЙ УСТАНОВКИ С
ВВЭР-1000/1200, ОБУСЛОВЛЕННЫХ ШТАТНЫМИ
ТЕХНОЛОГИЧЕСКИМИ ОПЕРАЦИЯМИ**

© 2022 Аркадов Геннадий Викторович¹, Трыкова Ирина Владимировна²,
Коцоев Константин Игоревич³

^{1,2,3}АО Научно-технический центр «Дианпром», Обнинск, Калужская обл., Россия

³Московский государственный технический университет им. Н.Э. Баумана, Москва, Россия

¹skrepka1964@gmail.com

²trykova@kvantprogramm.ru

³kotsoev@kvantprogramm.ru

Аннотация. В реакторных установках с водо-водяным энергетическим реактором не исключено появление в главном циркуляционном контуре свободных, слабозакрепленных и посторонних предметов. Эти предметы, перемещаясь в потоке теплоносителя, могут соударяться со внутренними стенками главного циркуляционного контура, что может привести к повреждению оборудования. Раннее обнаружение этих предметов позволит минимизировать повреждения и повысить уровень безопасности эксплуатации АЭС. С этой целью реакторная установка оснащается системой обнаружения свободных/слабозакрепленных предметов (СОСП). Основной проблемой СОСП является большое количество ложных тревог, возникающих вследствие регистрации шумов от штатного функционирования АЭС. В работе рассматривается применение алгоритмов кластеризации к сигналам СОСП, что позволяет значительно уменьшить число ложных тревог, поскольку установлено, что сигналы от срабатывания штатного оборудования отличаются большей степенью повторяемости. Тогда, «обучив» СОСП на некотором архиве данных, характеризующих штатное функционирование РУ, мы можем утверждать, что, если вновь поступивший сигнал попадает в один из кластеров, то он отражает штатное функционирование РУ, в то время как сигналы, не попавшие ни в один из кластеров, могут быть следствием появления свободного/слабозакрепленного предмета и данная ситуация требует незамедлительного реагирования персонала, эксплуатирующего АЭС. Данный подход позволяет значительно уменьшить количество выходной информации СОСП, снизить нагрузку на эксплуатирующий персонал, улучшить качество принимаемых решений и, соответственно, увеличить безопасность эксплуатации РУ в целом.

Ключевые слова: атомная электростанция, реакторная установка, акустические сигналы, кластеризация, большие данные, свободные и слабозакрепленные предметы, повреждение оборудования, безопасность эксплуатации, акустические аномалии, раннее обнаружение.

Для цитирования: Аркадов Г.В., Трыкова И.В., Коцоев К.И. Кластеризация акустических событий в главном циркуляционном контуре реакторной установки с ВВЭР 1000/1200, обусловленных штатными технологическими операциями // Глобальная ядерная безопасность. – 2022. – № 3(44). – С. 43-55. <http://dx.doi.org/10.26583/gns-2022-03-04>

Поступила в редакцию 30.05.2022

После доработки 29.07.2022

Принята к публикации 12.08.2022

Введение

Система обнаружения свободных и слабозакрепленных предметов (СОСП) является средством автоматического обнаружения аномальных акустических событий, регистрируемых акселерометрами, смонтированными на главном циркуляционном контуре реакторной установки (РУ) [1-2]. Основной целью СОСП является обнаружение на ранней стадии свободных и слабозакрепленных предметов в первом контуре. Эксплуатация системы на АЭС показала, что наряду со свободными и слабозакрепленными предметами, всплески акустического шума могут быть вызваны:

1) теплогидравлическими процессами (в том числе гидравлическими ударами, кавитацией, кипением теплоносителя (ТН), детонацией гремучей смеси);

2) штатными технологическими операциями или изменением состояния/режима работы агрегата какой-либо системы энергоблока (включая процедуры опробования и испытания оборудования);

3) вибрациями оборудования и трубопроводов главного циркуляционного контура (ГЦК).

Акустические эффекты также проявляются при тепловом перемещении оборудования в режимах разогрева/расхолаживания. Здесь возможны соударения элементов конструкции (в узлах крепления внутрикорпусных устройств, тепловыделяющих сборок (ТВС)), проявление эффекта «страгивания» подвижных опор парогенераторов (ПГ) и главных циркуляционных насосов (ГЦН) из состояния покоя. События, вызванные этими причинами, могут быть отнесены к ложным тревогам. Однако идентифицировать событие, вызванное технологическими причинами не всегда просто из-за сложности контролируемого оборудования, сложности систем, обеспечивающих его функционирование, сложности алгоритмов (технологии) их работы.

Технологические операции имеют, как правило, воспроизводимый «акустический портрет». Причем воспроизводятся не только диагностические признаки, но и собственно сигналы (вплоть до «поточечного» совпадения сигнала как функции времени). Например, срабатывание определенной арматуры дает один и тот же переходный процесс (акустическую аномалию). То есть, этот процесс можно отнести к детерминированным процессам.

Соударение свободного предмета всегда характеризуется наличием случайного (стохастического) компонента в сигнале датчика. Выявление стохастического и детерминированных компонентов сигнала возможно только статистическим путем. Регистрация кратных событий (последовательных во времени событий со схожими характеристиками), вероятнее всего, обусловлена технологическими причинами.

Таким образом, если классифицировать кратные события, которые имеют технологическую природу, количество ложных срабатываний СОСП значительно уменьшится.

Подходы к кластеризации событий

В идеализированном случае процесс классификации должен основываться на априорном знании источника акустической аномалии. В тестовом случае удар задается пользователем СОСП и, после его обнаружения и регистрации, оцениваются признаковые параметры (номер базового канала (БК), разность времен прихода, максимум среднеквадратичного значения, крутизна переднего фронта), которые объявляются образом (эталоном) данного класса. По возможности, перебираются последовательно все источники акустических эффектов. К таковым источникам, прежде всего, относится запорно-регулирующая арматура. Помимо такой тестовой классификации событий возможен и функциональный подход, когда в процессе нормального функционирования объекта, после того как зарегистрирован акустический

эффект, выявляется его источник (некоторое штатное управление объектом). Тестовую классификацию можно осуществить в режиме обязательного ежегодного опробования арматур систем безопасности. Она осуществляется в режиме «горячего» останова по утвержденной программе. Срабатывания арматур систем безопасности генерирует, как правило, мощные акустические сигналы. Однако потом, в режиме нормальной эксплуатации, данные события не воспроизводятся и практический эффект по снижению числа ложных тревог становится нулевым.

Функциональную классификацию, по возможности, необходимо проводить всегда, в течение всего срока эксплуатации СОСП. Она состоит в установлении «задним числом» истинного источника зарегистрированного удара. Именно по такому принципу российские и зарубежные разработчики СОСП предполагают обучение системы. Для этого целесообразно искать совпадения во времени некоторого дискретного штатного сигнала управления со временем обнаружения акустического события СОСП. К сожалению, эта простая идея не реализуется в силу технических проблем обмена информацией между различными системами автоматизированной системы управления технологических процессов (АСУ ТП).

Нисколько не отвергая тестовый и функциональный подходы к классификации, в данной работе реализован статистический подход. Огромное число зарегистрированных в базе данных СОСП ударов позволяет применить в известном смысле формальные методы – алгоритмы кластеризации.

Прежде всего, здесь дается ответ на вопрос: «Наблюдалось ли в точности такое же событие ранее?» Если ответ «да», то зарегистрированное событие – ложная тревога, т.к. основное предназначение СОСП – обнаружение свободных предметов. Они, как известно, обладают большим стохастическим компонентом.

В статистическом подходе, как правило, формализуются большое число классов. Не каждый полученный класс может быть физически обоснован. Например, нарушается принцип «одна арматура – один класс».

При дальнейшей эксплуатации и накоплении информации, классы могут укрупняться. При статистическом подходе достигается практический результат, состоящий в том, что уменьшается угроза «обнаружения» свободного предмета в тех случаях, когда такового предмета в действительности нет, т.е. уменьшается вероятность ложной тревоги при диагностировании (ошибки первого рода).

Обзор методов классификации и кластеризации, методы оценки качества разбиения данных приведены в [3].

Огромный опыт по применению кластерного анализа к описанию состояний оборудования АЭС приведен в [4]. Там же показана важность выбора информативных признаков для кластеризации, приведены методы визуализации кластеров, произведено сопоставление каждого кластера к определенному штатному режиму работы оборудования либо к неисправности (аномалии).

Исходные данные

В РУ ВВЭР-1000/1200 существуют пять объемов ГЦК, конструктивно локализирующих свободные предметы:

- верхняя часть объема горячего коллектора ПГ (4 шт.);
- нижний объем между шахтой внутрикорпусных устройств и корпусом реактора.

Датчики СОСП располагаются вблизи мест естественных ловушек свободных предметов (рис. 1) при циркуляции ТН как при прямом (днище реактора, горячий коллектор ПГ), так и при обратном токе ТН (холодный коллектор ПГ).

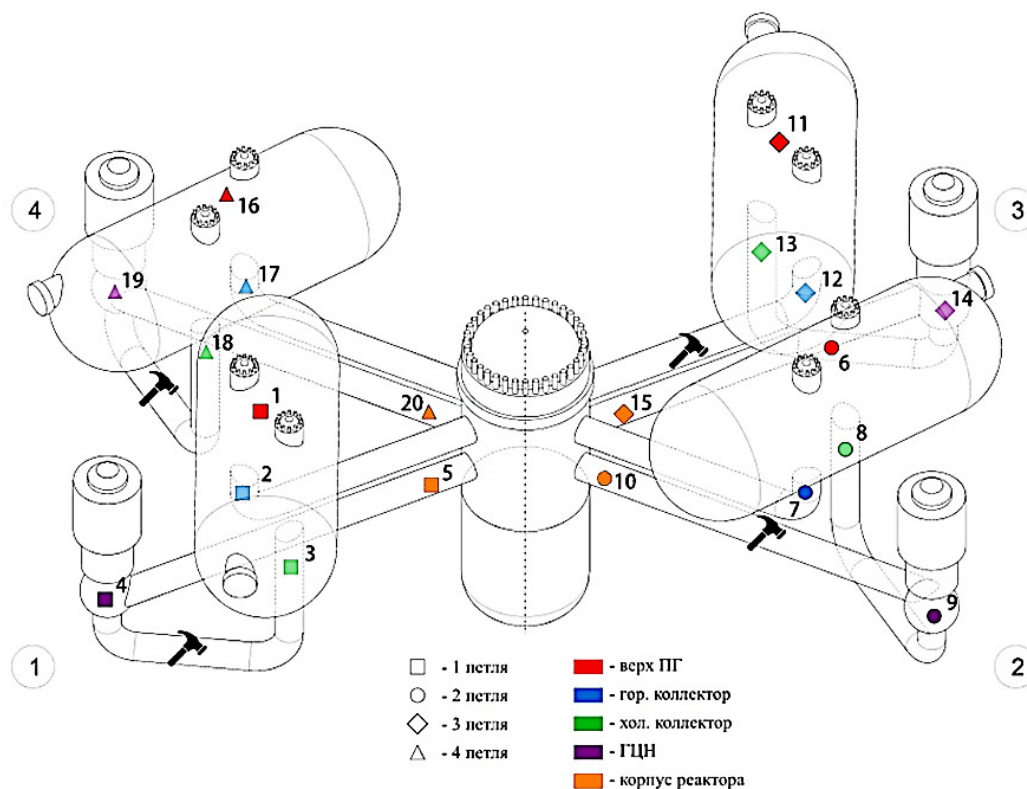


Рисунок 2 – Типовая схема расположения датчиков каналов обнаружения СОСП [Typical layout of detection channel sensors of Free and Loose Object Detection System]

В качестве исследуемого множества событий были взяты измерения, накопленные за 2021 г., одной из действующих СОСП РУ с ВВЭР-1200. Всего рассматривается порядка 11400 событий. Каждое событие переводится в признаковое пространство X . «Событие», регистрируемое СОСП, представляет собой запись сигналов со всех акустических датчиков одновременно, т.е. одно событие – это массив из 20 временных реализаций.

Временная реализация сигналов событий – это запись входного процесса, представляющая собой последовательность оцифрованных значений. Частота дискретизации при записи временных реализаций СОСП постоянна и равна 50 кГц. Длительность временных реализаций составляет 60 мс., что соответствует 3000 точек, и складывается из 20 мс. до превышения момента регистрации события – обнаружения факта превышения порога по первому каналу и 40 мс. после регистрации превышения.

Ключевым признаком события является номер базового канала (БК) – номер канала датчика, который первым зарегистрировал прохождение фронта волны от удара. Номер БК характеризует локализацию источника и максимум среднеквадратичного значения (СКЗ), характеризующего энергию акустического сигнала.

При этом СКЗ j -го сигнала события определяется по формуле (1):

$$r_{nj} = \sqrt{\frac{1}{k-1} \sum_{i=n}^{n+k} s_{ij}^2}, n = 0, 1, 2, \dots, 2900 \quad (1)$$

где k – окно усреднения для вычисления СКЗ (стандартно 100 точек, или 2 мс.);

s_{ij} – значение j -го сигнала в точке i .

Получаем временные реализации СКЗ сигнала длиной 2900 точек (из начальной длины реализации вычитается длина буфера), или 58 мс.

По СКЗ j -го сигнала определяется его «мощность» M_j : максимальное СКЗ акустического сигнала с поправкой на среднее значение СКЗ на участке фонового измерения (2):

$$M_j = \max_n(r_{nj}) - \frac{1}{N_f} \sum_{n=1}^{N_f} r_{nj}, \quad (2)$$

где N_f – длина участка фона.

Обозначим M_{ij}^{\max} – мощность j -го ($j=1,2,\dots,20$) сигнала i -го ($i=1,2,\dots,N$) события в выборке. Помимо значений мощности, в качестве признаков будем использовать нормированные значения той же самой мощности на максимальную, но внутри одного события (3):

$$M_{ij}^{\text{rel}} = \frac{M_{ij}}{\max_j(M_{ij})}. \quad (3)$$

Исторически сложилось, что диагностическими (классификационными) признаками регистрируемых СОСП сигналов событий являются (для фиксированного БК), рисунок 2:

- номера каналов трех следующих по времени прихода переднего фронта волны датчиков;
- время прихода переднего фронта БК;
- разность времен прихода переднего фронта между базовым датчиком и тремя другими;
- мощность сигналов первых четырех сигналов по времени прихода переднего фронта;
- нормированные значения мощности первых четырех сигналов;
- крутизна переднего фронта первых четырех сигналов.

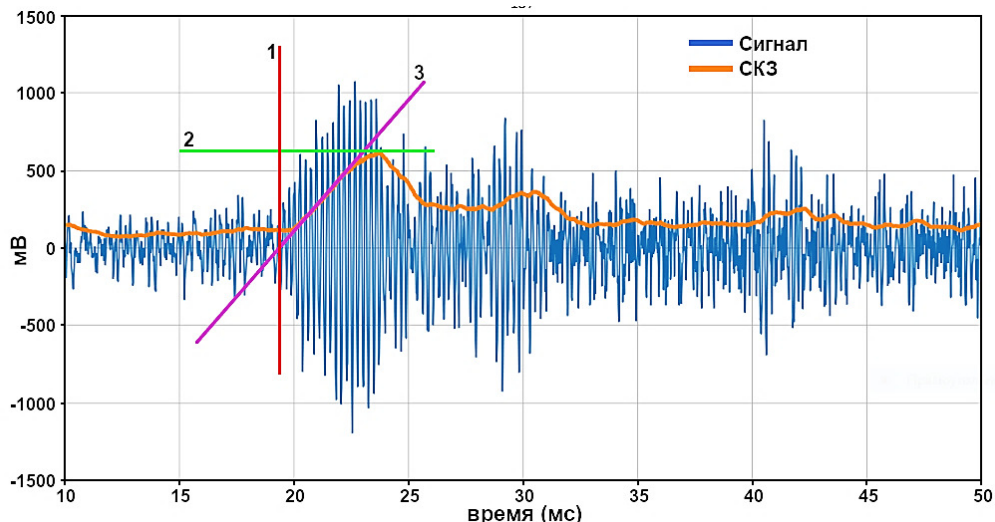


Рисунок 3 – Пример определения диагностических признаков: 1 – время прихода переднего фронта; 2 – максимум СКЗ; 3 – крутизна переднего фронта [Example of defining diagnostic features: 1 - leading edge arrival time; 2 - maximum RMS voltage; 3 - leading edge steepness]

Данное 19-мерное признаковое описание рассчитывается на основании временных реализаций сигналов всех 20 каналов, по 10 параметрам определения времени прихода. Подбор параметров определения времени прихода осуществляется экспертом, неаккуратно выставленные параметры влияют на очередность каналов по

времени прихода переднего фронта сигналов, что, в свою очередь, может исказить локализацию акустической аномалии.

Будучи разными по физическому смыслу, диагностические признаки сильно различаются между собой по абсолютным величинам. Дисбаланс между значениями признаков может вызвать неустойчивость работы алгоритмов кластеризации и ухудшить результаты обучения.

Для того чтобы нивелировать данные различия, предлагается вместо абсолютных значений этих признаков использовать их логарифмические значения и нормализацию. Для нормализации используем метод Мин-Макс [5]. Данный метод переводит измеренные значения в значения между 0.0 и 1.0. Нормированное значение признака X определяется как (4):

$$X_n = \frac{\lg X - \min(\lg X)}{\max(\lg X) - \min(\lg X)}. \quad (4)$$

Распределение событий по базовым каналам приведено на рисунке 3.

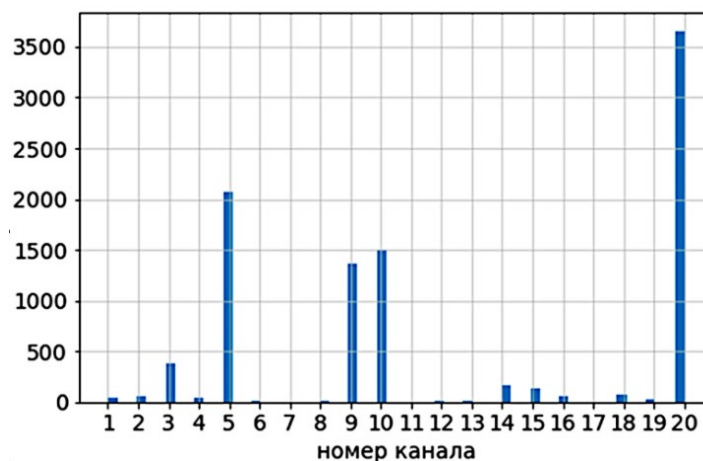


Рисунок 4 – Распределение событий по базовым каналам [Distribution of events to base channels]

Кластеризация K-means

Для кластеризации событий воспользуемся методом *Mini Batch K-means*, который был предложен в [5] как альтернативный алгоритм кластеризации для больших объемов данных. Преимуществом данного метода является снижение вычислительных затрат путем использования подвыборок меньшего фиксированного размера вместо всего набора исходных данных.

Метод основан на следующих интуитивных предположениях об оптимальных кластерах:

- «центром кластера» является арифметическое среднее всех точек кластера;
- каждая точка расположена ближе к центру своего кластера, чем к центрам соседних.

В данном методе решается задача минимизации целевой функции (5):

$$\min \sum_{x \in X} \|f(C, x) - x\|^2, \quad (5)$$

где C – матрица центров кластеров,

$f(C, x)$ – функция, возвращающая ближайший центр кластера ($c \in C$) к x на основании Евклидова расстояния.

Для оптимизации применяется метод стохастического градиентного спуска по подвыборке, которая имеет, как правило, меньший уровень шума, т.е. улучшается сходимость метода.

Первоначальная инициализация центров кластеров совершается по алгоритму *k-means++* [6]. Существенным недостатком метода *K-means* является то, что количество кластеров должно быть задано заранее.

Для кластеризации были отобраны сигналы по базовому каналу №5 (2073 шт.), которые были разбиты на 400 обобщенных кластеров, характеризующих мощность и характер акустического процесса. Результаты кластеризации приведены на рисунке 4.

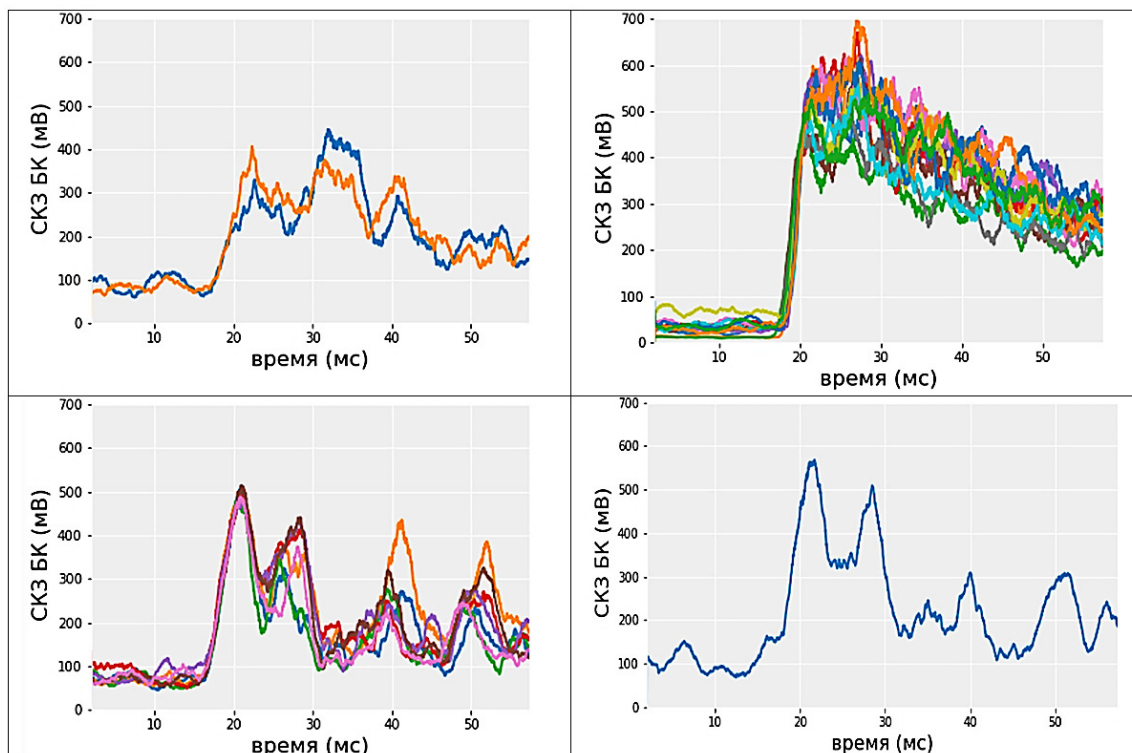


Рисунок 5 – Примеры кластеров, полученных при помощи алгоритма K-means, для БК №5 [Examples of clusters obtained using the K-means algorithm, for BC№5]

Кластеры, содержащие только один сигнал, могут быть порождены сигналами от свободных/слабозакрепленных предметов.

Недостатком данного подхода является то, что количество кластеров задается вручную и нет очевидной меры качества кластеров. Поэтому невозможно определить само качество кластеризации и процент некластеризованных событий.

Кластеризация DBSCAN

Так как количество классов априори неизвестно, то для кластеризации событий используем плотностной алгоритм *DBSCAN* (Density-based spatial clustering of applications with noise) [7], который не требует их задания при инициализации.

DBSCAN алгоритм рассматривает кластеры как области с высокой плотностью, разделенные областями с низкой плотностью. В связи с этим, с общей точки зрения, кластеры, найденные по *DBSCAN*, могут быть любой формы, в отличие от К-средних, которые предполагают кластеры выпуклой формы. Алгоритм вычисляет расстояния между соседними элементами и на основании этих расстояний выделяет кластеры.

Имея соответствующее описание события в численном признаковом пространстве, в качестве меры близости двух событий можно использовать Евклидово расстояние – геометрическое расстояние между событиями в \mathbb{R}^{19} пространстве признаков (6):

$$d(x_i, x_j) = \|x_i - x_j\|_2 = \sqrt{\sum_{n=1}^{19} |x_{i_n} - x_{j_n}|^2} \quad (6)$$

Центральным компонентом в *DBSCAN* является понятие корневых элементов, которые находятся в областях с высокой плотностью. Кластером является набор корневых объектов, которые расположены близко друг к другу (используя меру близости) и набор граничных объектов, которые близки к корневым, но сами не являются корневыми. Все события, которые не являются ни корневыми, ни граничными относятся к «шумовым», которые не принадлежат ни одному кластеру.

При инициализации алгоритма задаются два параметра: *MinPts* и ϵ . Данные параметры непосредственно определяют меру плотности и влияют на результат кластеризации. Формально, корневой объект e_i определяется как объект в ϵ -окрестности которого находится как минимум *MinPts* объектов, т.е. (7):

$$N_\epsilon(x_i) = \{x_j \in E \mid d(x_i, x_j) \leq \epsilon\} \geq \text{MinPts}. \quad (7)$$

Кластер представляет собой набор элементов, которые могут быть найдены рекурсивно, принимая образец корневого элемента, далее находятся все плотностно-достижимые соседи, которые определяются как корневые объекты, находятся все их соседи, и так далее. Кластер также имеет набор граничных объектов, которые являются соседними для корневого объекта, но сами не являются корневыми. Интуитивно эти объекты находятся на периферии кластера. Процесс завершается, когда ни к одному кластеру не может быть добавлено ни одного нового объекта.

Выбор параметров *MinPts* и ϵ основывается на специфике исходных данных [8]. Технологические операции, такие как срабатывания арматуры, отвечающие за безопасность, довольно редкие явления. Поэтому минимальное количество событий в одном кластере (*MinPts*) принимается равным двум. Данное количество минимального числа событий в кластере позволит кластеризовать как редкие, так и часто происходящие события, не оказав существенного влияния на качество кластеризации. Стоит отметить, что такие низкие значения *MinPts* в значительной степени увеличивают количество кластеров. Идея выбора оптимального значения ϵ заключается в нахождении «перегиба» на графике расстояний между двумя соседними событиями, упорядоченному по возрастанию. Из рисунка 5 видно, что оптимальное значение ϵ лежит в районе 0,2.

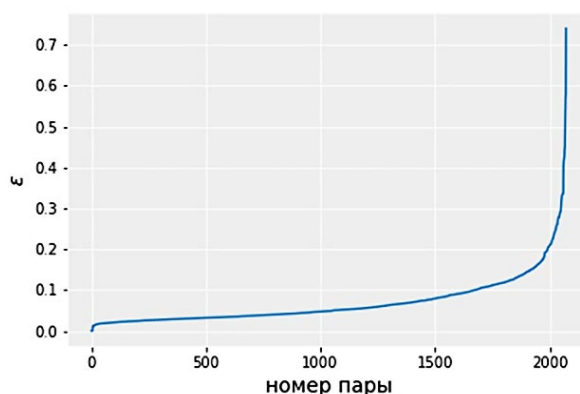


Рисунок 5 – Упорядоченное межпарное расстояние обучающей выборки кластеризации DBSCAN
[Ordered inter-pair distance of the DBSCAN clustering training sample]

В результате кластеризации для 2073 событий по БК №5 методом *DBSCAN* получается 57 кластеров (пример СКЗ сигналов разных кластеров показан на рисунке 6), которые покрывают ~ 81% всех зарегистрированных событий.

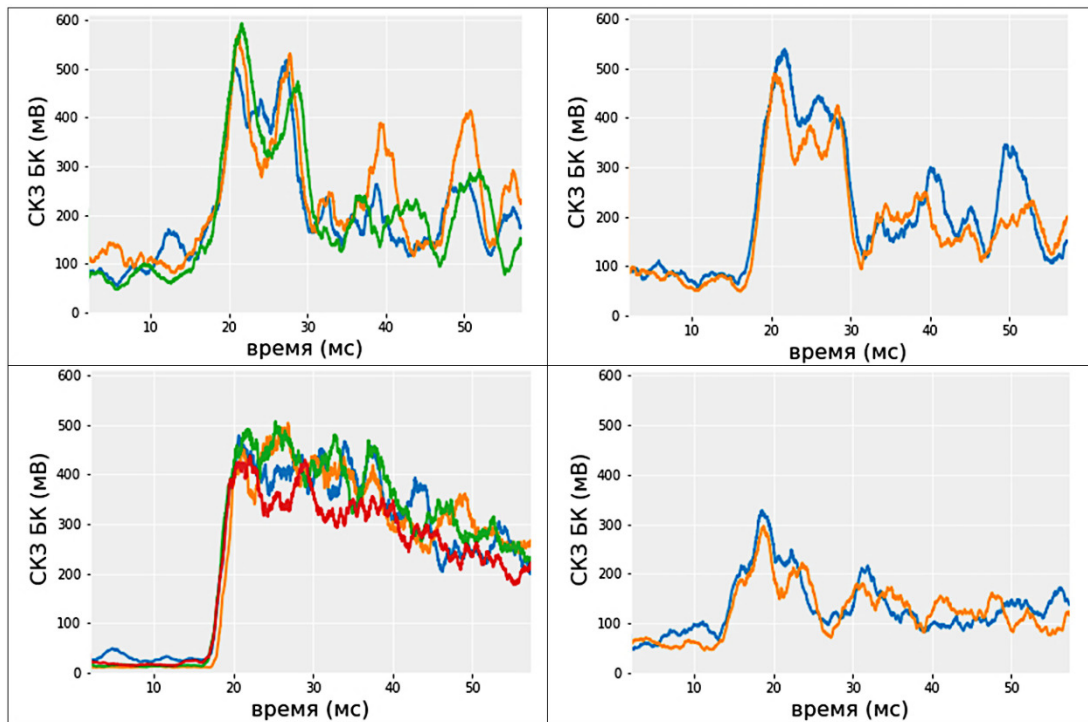


Рисунок 6 – Примеры кластеров DBSCAN по БК №5 [Examples of DBSCAN clusters by BC№5]

Иерархическая кластеризация

Иерархическая кластеризация [9-10] – совокупность алгоритмов упорядочивания данных, направленных на создание иерархии (дерева) вложенных кластеров. Дерево строится от листьев к корню. В начальный момент времени каждый объект содержится в собственном кластере. Далее происходит итеративный процесс слияния двух ближайших кластеров до тех пор, пока все кластеры не объединятся в один или не будет найдено необходимое число кластеров. Эффективность алгоритмов иерархической кластеризации в применении к идентификации режимов теплообмена была исследована в работе [4]. Подход, использующий для кластеризации участки временных рядов, разработан в [11].

Как уже отмечалось выше, штатные технологические операции имеют воспроизводимый «акустический портрет». Поэтому, воспользуемся похожестью формы сигнала, начиная от времени, когда сигнал был зарегистрирован и до некоторого момента, когда он начинает спадать. Для анализа берем 200 точек СКЗ сигнала по базовому каналу, откорректированного на уровень фона (см. формулу 2), начиная от времени прихода переднего фронта. Особенностью данного подхода является то, что нормализация Мин-Макс не производится.

На каждом шаге вычисляется расстояние между кластерами и пересчитывается расстояние между новыми кластерами. Расстояние между кластерами X и Y определяется через расстояние между объектами по методу «полной связи» (англ. *complete linkage*) [12] (8):

$$R(X,Y) = \max_{x \in X, y \in Y} \|x - y\|_2 = \max_{x \in X, y \in Y} \sqrt{\sum_{n=1}^{200} |x_n - y_n|^2} \quad (8)$$

Метод требует настройки длины фрагмента сигнала и порога метрики, отвечающей за принадлежность к кластеру.

Результат кластеризации для 1473 событий по БК №9 методом иерархической кластеризации показан на рисунке 7, получается 197 кластеров.

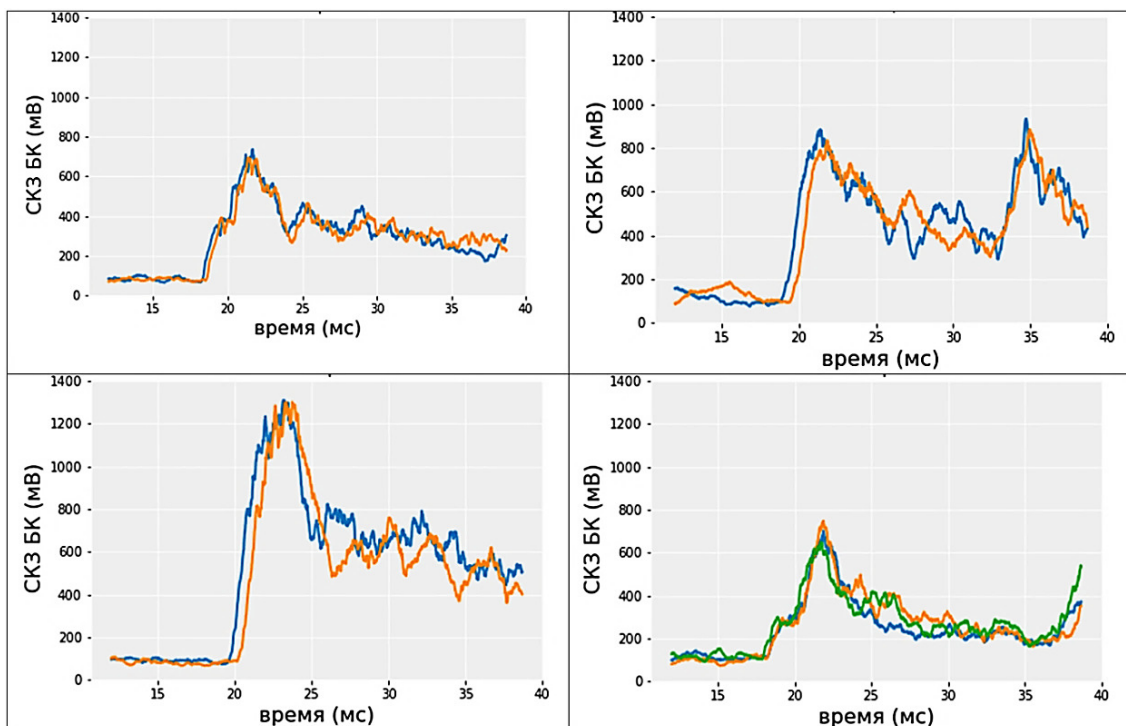


Рисунок 7 – Результаты иерархической кластеризации по БК №9 [Hierarchical clustering results for BC№9]

Выводы

В статье обобщен опыт применения кластерного анализа к данным, регистрируемым СОСП. Приведен метод выделения характерных признаков сигналов, необходимых для кластеризации событий. Вновь регистрируемые события проходят процедуру классификации, т.е. причисление события к ранее созданному классу (кластеру) по совпадению рассчитанных параметров сигналов. Данная процедура в значительной степени уменьшает количество ложных срабатываний СОСП.

Некластеризованные события относятся к «шумовым». Они связаны как со свободными или слабозакрепленными предметами, так и обусловлены ложными регистрациями СОСП, вызванными нестационарностью фонового шума, временной неисправностью некоторых каналов, заменой блоков предусилителей, влияющих на абсолютные значения сигналов и другими причинами.

Полученные кластеры не несут в себе информацию об источнике акустической аномалии, а только подтверждают факт ее повторения. Поэтому там, где это возможно, необходимо проведение дальнейшей обработки событий – уточнение причин их возникновения (перевод безымянных кластеров в область точных классов). Т.е. каждому кластеру должна быть поставлена в соответствие та технологическая операция, работой или изменением режима работы которой вызваны события этого кластера.

В данном случае решается обратная задача – установления связи «акустическая реакция – воздействие», которая производится по штатным сигналам АСУ ТП. Срабатывания арматуры приводит к воспроизводимым акустическим реакциям с большой степенью детерминированности [2]. Временные метки дискретных сигналов АСУ ТП, связанных с работой технологического оборудования, сопоставляются со временем регистрации акустической аномалии. Данная алгоритмическая процедура осложняется большим количеством регистрируемых технологических параметров и особенностями хранения исторической информации – разреживанием сигналов в системе верхнего блочного уровня (СВБУ – подсистема АСУ ТП). Однако, проведя процедуру предварительной кластеризации, достаточно определить источник только одного события из данного кластера. Так, например, на рисунке 8 показаны сигналы

событий двух кластеров, вызванных движением органов регулирования системы управления и защиты (ОР СУЗ).

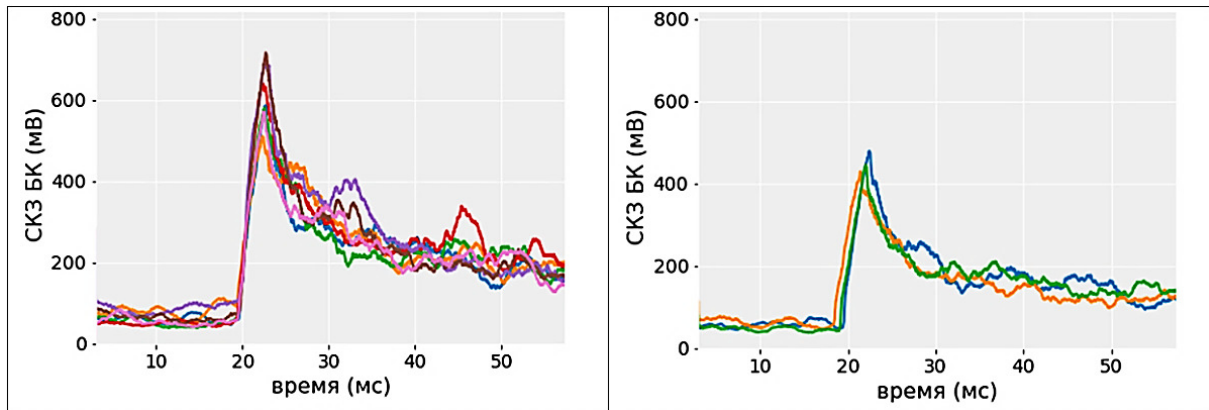


Рисунок 8 – Кластеры сигналов событий, вызванных движением ОР СУЗ [Clusters of events caused by control and protection system movements]

Помимо определенно детерминированных классов (классы с низким коэффициентом вариации), существует достаточное количество классов, характеризующихся высоким коэффициентом вариации по количественным признакам. Данные события вызваны теми технологическими операциями, которым присущ некоторый стохастический компонент, сильно влияющий на форму сигнала. К таким можно отнести срабатывания регулирующей арматуры, события, источником, которых является ГЦН и др.

Мерой качества кластеризации в нашем случае является «похожесть» СКЗ сигналов друг на друга в диапазоне времен от времени прихода переднего фронта сигнала до начала затухания (например, на рис.8 это диапазон от 18 мс до 30 мс). При этом различимость кластеров друг от друга значения не имеет, важен сам факт принадлежности сигнала какому-либо из кластеров. В качестве меры «компактности» кластера возьмем величину (9):

$$D = \max_k \frac{\sqrt{\sum_{i=N_1}^{N_2} (s_k(t_i) - s_0(t_i))^2}}{\sqrt{\sum_{i=N_1}^{N_2} (s_0(t_i))^2}}, \quad (9)$$

где: $s_0(t_i)$ – центроид кластера, $i = \{0, 1, 2, 3, \dots, 2900\}$;

$s_k(t_i)$ – произвольное СКЗ из кластера, $k = \{0, 1, 2, 3, \dots, M\}$;

N_1 и N_2 – точка прихода переднего фронта и точка начала затухания.

Центроид кластера найден как (10):

$$s_0(t_i) = \underset{k}{\text{mediana}} \{s_k(t_i)\}, \quad k = \{0, 1, 2, 3, \dots, M\}. \quad (10)$$

«Достоверными» будем считать те кластеры, для которых D меньше определенного порога, по рассмотренным выше данным был выбран порог $D = 0.25$, при этом количество «достоверных» кластеров составило выше 85% для всех рассмотренных методов кластеризации.

Благодарность

Авторы выражают благодарность научному руководителю данного исследования, заместителю директора по науке АО «НТЦ «Диапром», к.т.н. Павелко В.И. за оказанную помощь при написании настоящей статьи.

СПИСОК ИСТОЧНИКОВ

1. Центр диагностики «Диапром». Система обнаружения свободных предметов в главном циркуляционном контуре ВВЭР-1000 (СОСП). – URL: <http://www.diaprom.com/projects/?p=2> (дата обращения: 27.05.2022).
2. Аркадов, Г.В. Системы диагностирования ВВЭР / Г.В. Аркадов, В.И. Павелко, Б.М. Финкель. – Москва : Энергоатомиздат, 2010. – 391 с.
3. Айвазян, С.А. Прикладная статистика. Классификация и снижение размерности / С.А. Айвазян, В.М. Бухштабер, И.С. Енюков, Л.Д. Мешалкин. – Москва : Финансы и статистика, 1989. – 607 с.
4. Скоморохов, А.О. Многомерные статистические методы диагностики аномальных состояний ЯЭУ : диссертация на соискание учений степени доктора технических наук / А.О. Скоморохов. – Обнинск, 2011. – 302 с.
5. Sculley, D. Web-scale k-means clustering // Proceedings of the 19th international conference on World wide web. – USA, 26 April 2010, P. 1177-1178 (in English).
6. Arthur, D., Vassilvitskii, S. k-means++: the advantages of careful seeding // Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms. Society for Industrial and Applied Mathematics. – Philadelphia, PA, USA, 2007, P. 1027-1035(in English).
7. Ester, M., Kriegel, H.-P., Sander, J., Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise // Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining. – Portland, OR, 1996, P. 226-231 (in English).
8. Mullin, T. DBSCAN Parameter Estimation Using Python. – Jul 10, 2020. – URL: <https://medium.com/@tarammullin/dbscan-parameter-estimation-ff8330e3a3bd> [дата обращения 27.05.2022] (in English).
9. Pedro Pereira Rodrigues, João Gama, Joao Pedro Pedroso Hierarchical Clustering of Time-Series Data Streams // IEEE Transactions on Knowledge and Data Engineering. – 2008. –Vol. 20, No. 5, P. 615-627. – URL: https://www.researchgate.net/publication/3297889_Hierarchical_Clustering_of_Time-Series_Data_Streams [accessed 05/27/2022] (in English).
10. Жамбю, М. Иерархический кластер-анализ и соответствия / М. Жамбю – Москва : Финансы и статистика, 1988. – 345 с.
11. Amidon, A. How to Apply Hierarchical Clustering to Time Series / A. Amidon. Dec 9, 2020 – URL: <https://towardsdatascience.com/time-series-hierarchical-clustering-using-dynamic-time-warping-in-python-c8c9edf2fda5> (дата обращения: 27.05.2022).
12. Späth, H. Cluster Analysis Algorithms / H. Späth – Chichester: Ellis Horwood, 1980. – 226 p.

REFERENCES

- [1] Scientific and Technical Center «Diaprom». Systema obnaryzheniya svobodnykh predmetov v glavnom tzyrculyatzionnom konture VVER-1000(SOSP) [System for Detecting Loose Objects in the Main Circulation Circuit of PWR-1000 (SOSP)]. – URL: <http://www.diaprom.com/projects/?p=2> (accessed 05/27/2022) (in Russian).
- [2] Arkadov G.V., Pavelko V.I., Finkel B.M. Systemy diagnostirovaniya VVER [PWR Diagnostic Systems]. Moskva: Energoatomizdat [Moscow: Energoatomizdat Publishing House]. 2010. 391 p. (in Russian).
- [3] Ayvazyan S.A., Bukhshtaber V.M., Enyukov I.S., Meshalkin L.D. Prikladnaya statistika. Klassifikatsiya I snizhenie razmernosti [Applied Statistics. Classification and Dimensionality Reduction]. Moskva: Finansy I statistika [Moscow: Finance and Statistics, Publishing House]. 1989. 607 p. (in Russian).
- [4] Skomorokhov A.O. Mnogomerniye statisticheskiye metody diagnostiki anomalnykh sostoyaniy YAYEU: dissertatsiya na soiskaniye ucheniy stepeni doktora tekhnicheskikh nauk [Multidimensional Statistical Methods for Diagnosing Anomalous States of Nuclear Power Plants: Thesis for the Degree of Doctor of Technical Sciences]. Obninsk, 2011. 302 p. (in Russian).
- [5] Sculley D. Web-Scale K-Means Clustering // Proceedings of the 19th International Conference on World Wide Web. – USA, 26 April 2010, – P. 1177-1178 (in English).

- [6] Arthur D., Vassilvitskii S. K-Means++: the Advantages of Careful Seeding // Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms. Society for Industrial and Applied Mathematics. – Philadelphia, PA, USA, 2007, P. 1027-1035 (in English).
- [7] Ester M., Kriegel H.-P., Sander J., Xu X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise // Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining. – Portland, OR, 1996, P. 226-231 (in English).
- [8] Mullin T. DBSCAN Parameter Estimation Using Python / T. Mullin; Jul 10, 2020. – URL: <https://medium.com/@tarammullin/dbscan-parameter-estimation-ff8330e3a3bd> (accessed 27/05/2022) (in English).
- [9] Pedro Pereira Rodrigues, João Gama, Joao Pedro Pedroso Hierarchical Clustering of Time-Series Data Streams // IEEE Transactions on Knowledge and Data Engineering. – 2008. – Vol. 20, No. 5, P. 615-627. – URL: https://www.researchgate.net/publication/3297889_Hierarchical_Clustering_of_Time-Series_Data_Streams [reference date 27.05.2022] (in English).
- [10] Jambyu M. Ierarhicheskiy klaster-analiz I sootvetstviya [Hierarchical Cluster Analysis and Correspondences]. Moskva: Finansy I statistika [Moscow: Finance and Statistics, Publishing House], – 1988. – 345 p. (in Russian).
- [11] Amidon A. How to Apply Hierarchical Clustering to Time Series,. Dec 9, 2020. – URL: <https://towardsdatascience.com/time-series-hierarchical-clustering-using-dynamic-time-warping-in-python-c8c9edf2fda5> (accessed 05/27/2022) (in English).
- [12] Späth H. Cluster Analysis Algorithms – Chichester: Ellis Horwood, 1980. 226 p. (in English).

Clustering of Acoustic Events in Main Circulation Circuit of WWER 1000/1200 Reactor Facility Caused by Normal Technological Operations

Gennadiy V. Arkadov¹, Irina V. Trykova², Konstantin I. Kotsoev³

^{1,2,3}*Joint Stock Company Scientific and Technical Center «Diaprom», Koroleva str., 6, Obninsk, Kaluga region, Russia 249031*

³*Bauman Moscow State Technical University, 2 Bauman str., 5/1, Moscow, Russia 105005*

¹*skrepka1964@gmail.com*

²*trykova@kvantprogramm.ru*

³*kotsoev@kvantprogramm.ru*

Abstract. The appearance of free, weakly fixed and foreign objects in the main circulation circuit is not ruled out in reactor plants with a pressurized water power reactor. These objects, moving in the coolant flow, can collide with the inner walls of the main circulation circuit, which can lead to equipment damage. Early detection of these objects will minimize damage and improve the safety of NPP operation. The reactor plant is equipped with a system for detecting loose/weakly fixed objects for this purpose. The main problem is a large number of false alarms arising from the registration of noise from the normal operation of the NPP. The paper considers the application of clustering algorithms to signals of the system for detecting loose/weakly fixed objects, which can significantly reduce the number of false alarms as it has been established that signals from the operation of standard equipment are highly repeatable. Then, having «trained» the system on a certain archive of data characterizing the regular functioning of the NPP, we can state that if the newly received signal falls into one of the clusters, then it reflects the normal functioning of the NPP, while the signals do not that fell into any of the clusters may be the result of the appearance of a loose / loosely fixed object, and this situation requires an immediate response from the personnel operating the NPP. This approach makes it possible to reduce the amount of the system for detecting loose/weakly fixed objects output information significantly, reduce the load on the operating personnel, improve the quality of decisions made and, accordingly, increase the safety of operation of the reactor plant as a whole.

Keywords: nuclear power plant, reactor plant, acoustic signals, clustering, big data, loose and weakly fixed objects, equipment damage, operational safety, acoustic anomalies, early detection.

For citation: Arkadov G.V., Trykova I.V., Kotsoev K.I. Clustering of acoustic events in the main circulation circuit of a WWER 1000/1200 reactor facility caused by normal technological operations // Global nuclear safety. 2022. Vol. 3(44). P. 43-55. <http://dx.doi.org/10.26583/gns-2022-03-04>