

ЭКСПЛУАТАЦИЯ ОБЪЕКТОВ
АТОМНОЙ ОТРАСЛИ
OPERATION OF FACILITIES
NUCLEAR INDUSTRY

УДК 621.039.5
DOI 10.26583/gns-2023-01-05
EDN MXCNRN

**ПРИМЕНЕНИЕ ИЕРАРХИЧЕСКОЙ АГЛОМЕРАТИВНОЙ
КЛАСТЕРИЗАЦИИ ДЛЯ СОЗДАНИЯ БИБЛИОТЕКИ БАЗОВЫХ
КЛАССОВ АКУСТИЧЕСКИХ СОБЫТИЙ РЕАКТОРНОЙ
УСТАНОВКИ С ВВЭР-1000/1200**

© 2022 Геннадий Викторович Аркадов¹, Ирина Владимировна Трыкова²,
Денис Валерьевич Звягинцев³, Константин Игоревич Коцоев⁴

^{1,2,3,4} Акционерное общество «Научно-технический центр «Дианпром», Обнинск, Калужская обл., Россия

⁴ Московский государственный технический университет им. Н.Э. Баумана, Москва, Россия

¹ skrepka1964@gmail.com

² trykova@kvantprogramm.ru

³ zvyagincev@diaprom.ru

⁴ kotsoev@kvantprogramm.ru

Аннотация. В реакторных установках с водо-водяным энергетическим реактором (ВВЭР) возможно появление в главном циркуляционном контуре (ГЦК) свободных, слабозакрепленных и посторонних предметов, представляющих угрозу целостности оборудования и безопасности РУ. С целью раннего обнаружения этих предметов РУ оснащается системой обнаружения свободных/слабозакрепленных предметов (СОСП). Помимо обнаружения свободных/слабозакрепленных предметов, в функции СОСП входит классификация зарегистрированных событий. Возможность применения алгоритма классификации основывается на том, что сигналы от срабатывания штатного оборудования отличаются большей степенью повторяемости, даже при наличии шума, в то время как свободный предмет характеризуется большим стохастическим компонентом и для него не может быть сформирован свой детерминированный класс. Классификация уменьшает число ложных тревог, позволяя выделить сигналы от штатных операций, при этом сигналы от одного процесса должны быть отнесены к одному классу. Идея статьи заключается в том, чтобы «обучить» СОСП на некотором архиве данных, характеризующих штатное функционирование РУ, создать библиотеку «базовых» классов и выставить границы каждого класса так, чтобы, с одной стороны, учесть возможную вариабельность параметров сигналов вследствие шума. Определив базовые классы, мы можем утверждать, что, если вновь поступивший сигнал попадает в один из классов, то он отражает штатное функционирование РУ, в то время как сигналы, не попавшие ни в один из классов, могут быть следствием появления свободного/слабозакрепленного предмета. В статье проанализировано множество событий, накопленных в архиве одной из действующих СОСП. Произведена их кластеризация, в результате чего выделены классы событий, соответствующие штатным технологическим операциям. Для каждого класса вычислен центр класса и допустимые границы отклонений от центра. Все полученные центры классов представляют собой эталоны, по которым СОСП в реальном времени либо классифицирует вновь обнаруженное событие, либо характеризует его как «неклассифицированное».

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

Для цитирования: Аркадов Г.В., Трыкова И.В., Звягинцев Д.В., Коцоев К.И. Применение иерархической агломеративной кластеризации для создания библиотеки базовых классов акустических событий реакторной установки с ВВЭР-1000/1200 // Глобальная ядерная безопасность. – 2023. – № 1(46). – С. 54–66 <http://dx.doi.org/10.26583/gns-2023-01-05>.

Поступила в редакцию 23.11.2022
После доработки 31.01.2023
Принята к публикации 13.02.2023

Введение

СОСП является средством автоматического обнаружения акустических событий, регистрируемых акселерометрами, смонтированными на ГЦК РУ [1-4]. Основной целью СОСП является обнаружение на ранней стадии свободных и слабозакрепленных предметов в первом контуре. Но, помимо этой прямой задачи, есть еще ряд важных вопросов, которые необходимо решать в рамках диагностирования РУ.

Одной из таких задач является классификация акустических событий, имеющих технологическое происхождение и соотнесение их с реальными технологическими процессами. Технологические операции имеют, как правило, воспроизводимый «акустический портрет». Например, срабатывание определенной арматуры дает один и тот же переходный процесс. То есть этот процесс можно отнести к детерминированным процессам.

Эксплуатация СОСП показала, что детерминированные акустические события могут быть зарегистрированы вследствие:

- 1) теплогидравлических процессов;
- 2) штатных технологических операций или изменения состояния/режима работы энергоблока (включая процедуры опробования и испытания оборудования);
- 3) вибраций оборудования и трубопроводов главного циркуляционного контура (ГЦК);
- 4) теплового перемещения оборудования в режимах разогрева/расхолаживания;
- 5) срабатывания запорно-регулирующей арматуры и роторного оборудования.

Однако идентифицировать событие, вызванное технологическими причинами, не всегда просто из-за сложности контролируемого оборудования, сложности систем, обеспечивающих его функционирование, огромного количества процессов, происходящих одновременно, достаточно высокого уровня шумов.

Одним из способов идентификации событий является составление библиотеки классов с заданными возможными отклонениями в диагностических признаках.

Классы таких событий называются «базовыми», они получаются методом кластеризации архивов данных, накопленных за определенный период функционирования РУ. Кластеры определяются на основе событий, у которых многократно повторяются характеристики (так называемые «кратные» события). Далее для каждого кластера определяется центр и допустимое отклонение от центра. Параметры каждого такого кластера участвуют в формировании соответствующего базового класса и задаются при настройке СОСП для конкретного энергоблока.

При функционировании СОСП в режиме реального времени, для каждого вновь зарегистрированного события вычисляются соответствующие признаки. Классификация состоит в покомпонентном сравнении текущего вектора признаков со всеми «базовыми» векторами.

Если все компоненты текущего вектора совпали со всеми компонентами некоторого базового вектора (с учетом допустимого отклонения), то считается, что классификация успешно завершена. Пользователю СОСП сообщается, что обнаруженное событие принадлежит такому-то классу, соответствующему совпавшему образу-эталону. Если текущий вектор не совпал ни с одним базовым, то пользователю сообщается, что обнаруженное событие не классифицировано, классификация откладывается на отсроченный анализ (и возможно, это сигнал от свободного/слабозакрепленного предмета).

Данные для кластеризации

В качестве исследуемого множества событий были взяты архивы, накопленные одной из действующих СОСП РУ с ВВЭР-1200. Для кластеризации использовались 3967 событий.

Датчики измерительных каналов СОСП установлены на горячих и холодных коллекторах парогенераторов (ПГ), на верхней части корпуса ПГ, на холодных трубопроводах, а также на ГЦН [2]. «Событие», регистрируемое СОСП, представляет собой запись сигналов со всех акустических датчиков синхронно в одном и том же временном интервале длиной 60 мс, т.е. одно событие – это массив из 20 временных реализаций (рис. 1). Временные реализации сигналов события – это записи входных процессов, представляющие 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. записи фоновых шумов и 40 мс. записи после регистрации превышения установленного порога.

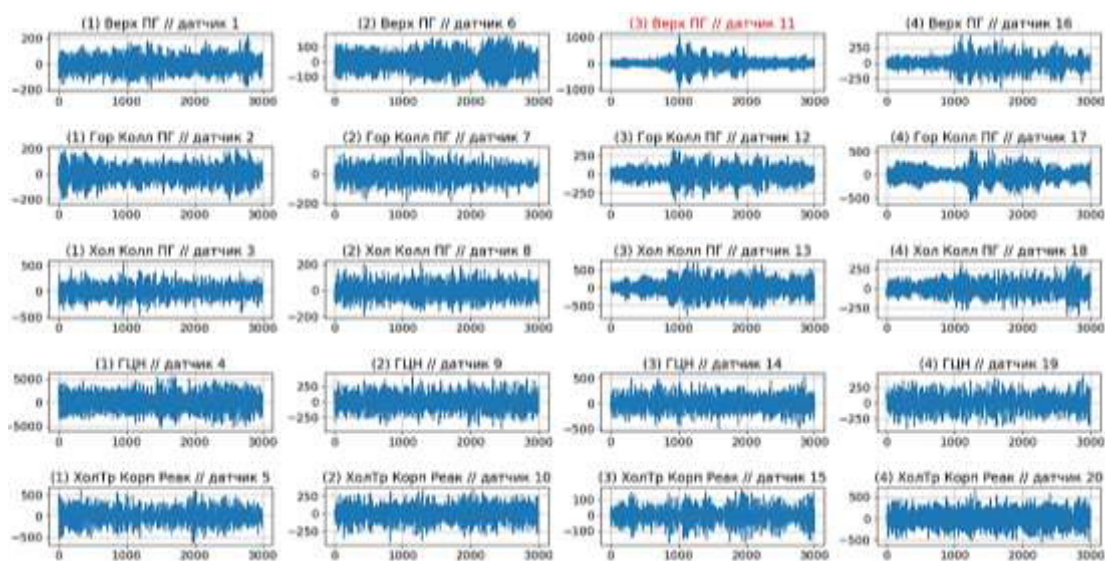


Рисунок 1 – Событие по петле 3, БК №11 [Loop 3 activity, BC No. 11]

Каждое событие переводится в признаковое пространство векторов. Диагностические (классификационные) векторы признаков регистрируемых СОСП событий вычисляются на основе временных реализаций отдельных сигналов события (рис. 2).

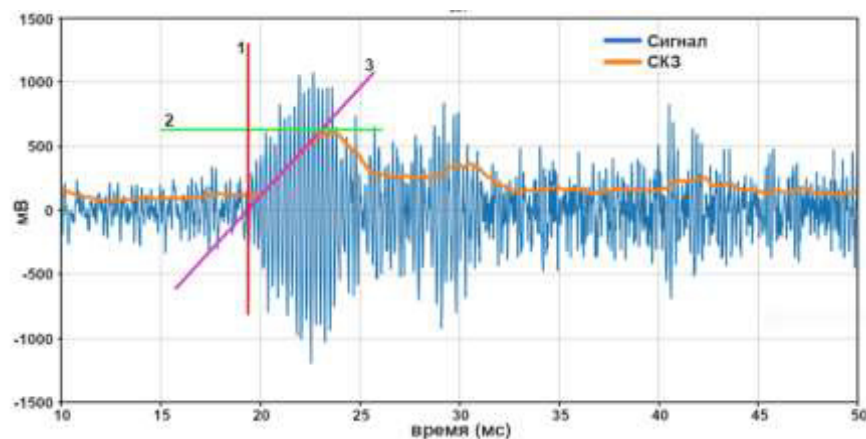


Рисунок 2 – Пример определения диагностических признаков: 1 – время прихода переднего фронта; 2 – максимум СКЗ; 3 – крутизна переднего фронта [Example of defining diagnostic features: 1 – arrival time of the leading edge; 2 – maximum RMS value; 3 – steepness of the leading edge]

После регистрации очередного события, из 20-ти временных реализаций, составляющих это событие, вычисляется среднеквадратичное значение (СКЗ) и текущий вектор, по которому производится классификация (рис. 3). Он состоит из 16-ти компонентов, которые, в свою очередь, собраны в 4 группы (4×4). Первая четвёрка – параметры первого сигнала (по времени прихода, то есть базового канала – БК), вторая – второго и т.д.: «N» – номер канала, «ВП N» – время прихода для канала N, «Кр N» – крутизна фронта, «СКЗ N» – максимум СКЗ для канала N.



Рисунок 3 – Вектор признакового пространства [Feature space vector]

Данное 16-мерное признаковое описание рассчитывается на основании 10 параметров 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.

Ключевым признаком события является номер БК – номер канала датчика, который первым зарегистрировал прохождение фронта волны от удара. Номер БК характеризует локализацию источника [5], максимум СКЗ характеризует энергию акустического сигнала.

Следует отметить, что кластеризация по всем 16-ти значениям используется только для сигналов очень большой амплитуды, таких, например, как сигнал от импульсных молотков [6], используемых для проверки работоспособности измерительных каналов СОСП. Для кластеризации сигналов от штатных операций используются значения для каналов N1 и N2, значения для каналов N3 и N4 – игнорируются. Данный подход обусловлен тем, что на РУ во время штатного функционирования возникает большое количество шумов различной природы и определить время прихода уже третьего канала (после БК) – затруднительно.

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

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

где k – окно усреднения для вычисления СКЗ;

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

Алгоритм получения первичных кластеров

Перед применением агломеративной кластеризации, необходимо выделить первичные кластеры. В теории распознавания образов выделяют три подхода к задаче синтеза первичных образов [7]:

- 1) принцип сравнения с эталоном;
- 2) принцип кластеризации;
- 3) принцип общности свойств.

Несколько алгоритмов кластеризации (принцип кластеризации) событий СОСП, были рассмотрены нами в работе «Кластеризация акустических событий в главном циркуляционном контуре реакторной установки с ВВЭР-1000/1200, обусловленных штатными технологическими операциями» [8]. В настоящей статье предпочтение было отдано детерминированному алгоритму (принцип сравнения с эталоном). Этот алгоритм применим для малого, вплоть до двух событий, набора данных (в отличие от приведенных в [8]). В основу положен метод, аналогичный алгоритму К внутригрупповых средних [7]:

1) для каждого компонента вектора признаков задаются минимально разумные пороги «от» и «до», далее в качестве центра первого кластера берется событие из архива, выделяются все события, попадающие в заданные границы, и уточняется центр кластера;

2) далее происходит итеративное уточнение центров кластеров, с целью минимизации суммы квадратов внутригрупповых расстояний. Внутри каждого, полученного на шаге 1 кластера, рассчитываются новые центры кластеров X_k . При этом предполагается, что сумма квадратов расстояний $\|x - X_k\|$ по всем x из k -го кластера должна быть минимальной. В этом случае центр X_k является выборочным средним для данного кластера и вычисляется по формуле (2):

$$X_k = \frac{1}{N_k} \sum_{j=1}^{N_k} x_j, \quad (2)$$

где N_k – количество образов x_j в k -м кластере;

3) если рассчитанные на шаге 2 новые центры совпадают с центрами, рассчитанными на предыдущей итерации (или отличаются от них не более чем на заданную величину ε), процесс заканчивается. В противном случае переходим к шагу 2.

Доказательство сходимости метода К внутригрупповых средних было получено в работе [9].

Недостатком всех вышеописанных алгоритмов является то, что, как правило, формализуется большое число кластеров (вследствие большого уровня шума, вариабельности параметров и неточностей определения компонент вектора признаков), что приводит к низкому проценту кластеризации и ограниченной применимости полученных классов для дальнейшего анализа акустических событий в режиме реального времени.

При этом одному и тому же физическому процессу ставится в соответствие несколько различных кластеров (т.е. нарушается принцип «один процесс – один класс»), что приводит к ухудшению быстродействия СОСП и сложностям при идентификации события по его классу. Также нередко возникает ситуация, когда априори известно, какие из событий СОСП заведомо относятся к одному классу, и требуется уточнить границы этого класса.

Например, на рисунке 4 показаны 2 различных кластера, которые относятся к одному и тому же событию (удары импульсных молотков), но при этом, формально различны (различие в номере четвертого канала, которое для каналов 15 и 20 не является существенным, но требует уточнения границ кластера).

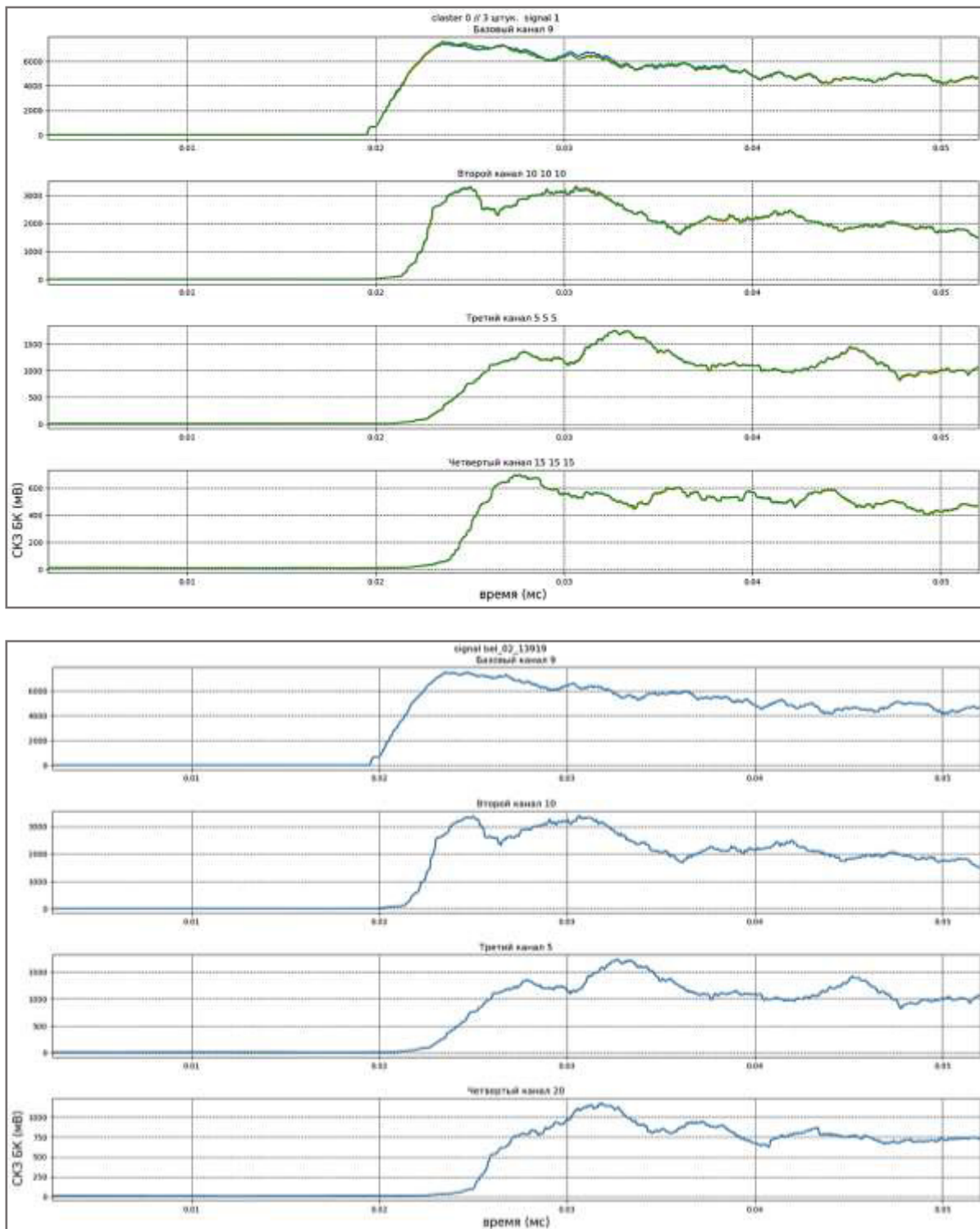


Рисунок 4 – События одного кластера, различающиеся по четвертому каналу [Activities of the same cluster, differentiated by fourth channel]

Для решения задачи уменьшения количества кластеров и повышения процента кластеризованных событий был применен алгоритм иерархической агломеративной кластеризации [10, 11]. При таком подходе наиболее близкие друг к другу центры первичных кластеров объединяются в один кластер, после чего уточняется центр нового, объединенного кластера и его допустимые границы. При этом оператор имеет возможность вручную задать максимальное отклонение от центра кластера (если он имеет некоторую априорную информацию о допустимом разбросе данных внутри кластера, что позволяет контролировать укрупнение кластеров).

Алгоритм иерархической агломеративной кластеризации

Идея иерархической агломеративной кластеризации основывается на последовательном объединении индивидуальных объектов или их кластеров во все более крупные кластеры. Результат работы алгоритма кластеризации представляет собой древовидную структуру или дендрограмму [12]. Исходя из дендрограммы определяется оптимальное количество кластеров.

Дендрограмма представляет собой визуальное представление меры расстояния между отдельными кластерами и визуально представляет каждый шаг процесса последовательного укрупнения кластеров. На двумерном графике по горизонтальной оси откладываются номера всех кластеров, а по вертикальной – расстояние между ними (рис. 5).

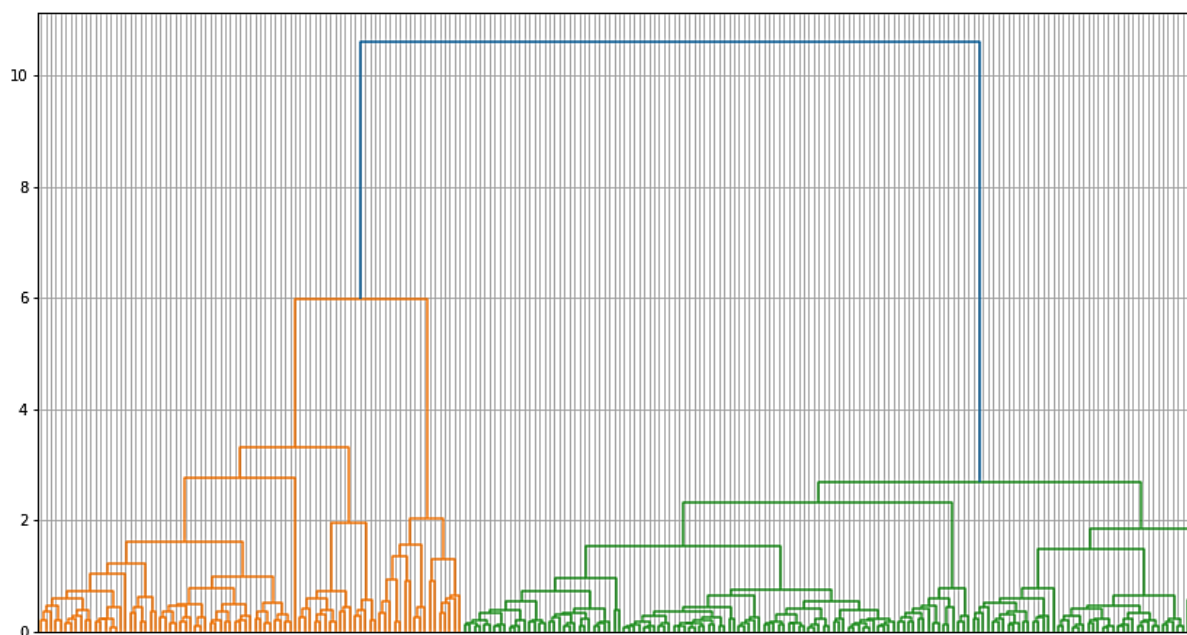


Рисунок 5 – Дендрограмма кластеров по БК 12 [Dendrogramme of clusters by BC 12]

При этом существует несколько методов вычисления расстояния между отдельными кластерами:

1) метод одиночной связи (single linkage) – «метод ближайшего соседа». Расстояние $D(A, B)$ между двумя кластерами A, B полагается равным минимальному расстоянию между двумя элементами из разных кластеров, выражение (3):

$$D(A, B) = \min \{d(a, b) : a \in A, b \in B\}, \quad (3)$$

где $d(a, b)$ – расстояние между элементами кластеров A и B ;

2) метод полной связи (complete linkage) также известен как «метод дальнего соседа». Расстояние между двумя кластерами полагается равным максимальному расстоянию между двумя элементами из разных кластеров, выражение (4):

$$D(A, B) = \max \{d(a, b) : a \in A, b \in B\}, \quad (4)$$

где $d(a, b)$ – расстояние между элементами кластеров A и B ;

3) метод средней связи (pair-group method using arithmetic mean) – расстояние между двумя кластерами полагается равным среднему расстоянию между элементами этих кластеров, выражение (5):

$$D(A, B) = \frac{1}{|A| \cdot |B|} \sum_{a \in A} \sum_{b \in B} d(a, b), \quad (5)$$

где $d(a, b)$ – расстояние между элементами кластеров A и B , а $|A|$ и $|B|$ – мощности кластеров;

4) центроидный метод (pair-group method using the centroid average) – расстояние между кластерами полагается равным расстоянию между их центрами, выражение (6) [13]:

$$D(A, B) = \|c_A - c_B\|, \quad (6)$$

где c_A и c_B – центроиды кластеров A и B , соответственно;

5) метод Уорда (Ward's method). В отличие от других методов кластерного анализа, для оценки расстояний между кластерами здесь используются методы дисперсионного анализа. В качестве расстояния между кластерами берётся прирост суммы квадратов расстояний объектов до центра кластера, получаемого в результате их объединения, выражение (7) [14]:

$$D(A, B) = \sum_i (x_i - \bar{x})^2 - \sum_{x_i \in A} (x_i - \bar{a})^2 - \sum_{x_i \in B} (x_i - \bar{b})^2. \quad (7)$$

На каждом шаге алгоритма объединяются такие два кластера, которые приводят к минимальному увеличению дисперсии. Этот метод применяется для задач с близко расположенными кластерами.

В данной работе использовался метод Уорда, он показал наибольшую устойчивость по отношению к выбросам данных и наилучшие интуитивно понятные результаты, исходя из специфики данных (зашумленность и большое количество близко расположенных кластеров).

Поскольку одна из основных функций СОСП – обнаружение свободных и слабозакрепленных предметов, отдельно следует остановиться на возможности ошибок второго рода («пропуск цели») – ложной классификации свободных предметов как «штатных технологических операций».

В качестве целевого множества для агломеративной кластеризации были использованы уже найденные на первичном этапе центры кластеров, это позволило работать именно с уже обнаруженными кластерами, что исключает ошибочную кластеризацию сигналов от возможных свободных предметов в качестве сигналов от технологических операций.

При определении границ объединенных кластеров также использовалось предположение о связности множества допустимых значений внутри одного кластера (т.е. все значения из интервала «от» и «до» принадлежат кластеру). Это позволило существенно увеличить долю кластеризованных событий при одновременном уменьшении количества кластеров.

Пользователь имеет возможность выбора из двух стратегий, исходя из конечной цели кластеризации:

1) ориентируясь на дендрограмму, установить точное количество кластеров – режим нужен при условии, что доступна априорная информация о количестве кластеров, как, например, показано на рисунке 4. В этом случае границы кластеров вычисляются исходя из имеющихся данных;

2) установить максимально допустимое отклонение от центров кластеров, количество кластеров при этом вычисляется автоматически. Границы кластеров также вычисляются исходя из имеющихся данных.

Ориентируясь на визуальное представление кластеров, оператор может принять дальнейшее решение об увеличении (или уменьшении) отклонения от центров.

В таблице 1 показаны границы кластеров, найденные методом агломеративной кластеризации, для различных предельных отклонений от центров. Первая строка содержит результаты первичной кластеризации, без использования агломеративной.

Таблица 1 – Результаты кластеризации с разными пороговыми значениями [Clustering results with texture thresholds]

Отклонение от центров, %	БК11 (1715 событий)	БК12 (276 событий)	БК16 (418 событий)
Δ ВП1=5% Δ Кр1=10% Δ СК31=10% Δ ВП2=5% Δ Кр2=10% Δ СК32=10%	кластеров: 207 кластеризовано: 46.4%	кластеров: 33 кластеризовано: 29.3%	кластеров: 64 кластеризовано: 48.6%
Δ ВП1=10% Δ Кр1=30% Δ СК31=30% Δ ВП2=10% Δ Кр2=30% Δ СК32=30%	кластеров: 143 кластеризовано: 56.6%	кластеров: 21 кластеризовано: 40.2%	кластеров: 38 кластеризовано: 60.5%
Δ ВП1=20% Δ Кр1=40% Δ СК31=40% Δ ВП2=20% Δ Кр2=40% Δ СК32=40%	кластеров: 62 кластеризовано: 61.8%	кластеров: 21 кластеризовано: 40.2%	кластеров: 22 кластеризовано: 66.3%

Пороги каждого кластера вычисляются индивидуально, если кластер «узкий» изначально – он таким и остается (табл. 2).

Таблица 2 – Пример определения центров и пороговых значений кластеров [An example of determining the centers and threshold values of clusters]

№ кластера	N1	ВП1	Кр1	СК31	N2	ВП2	Кр2	СК32
1	11	17.84	89.01	393.86	12	18.69	36.3	177.9
Δ 1, %	0.0	6.28	16.58	13.72	0.0	8.00	22.75	19.56
2	11	15.32	33.98	405.23	12	27.14	38.76	164.61
Δ 2, %	0.0	5.34	11.45	11.00	0.0	6.05	10.27	10.96

На рисунке 6 показан один из полученных кластеров.

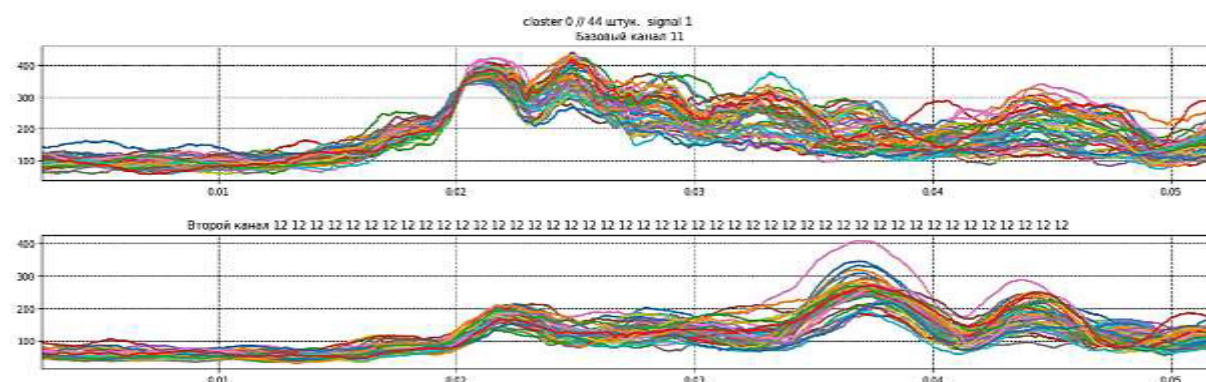


Рисунок 6 – Кластер по БК 11 [BC 11 cluster]

На рисунках 7, 8 приведено распределение количества кластеров с заданным числом элементов, до и после агломеративной кластеризации, соответственно. Видно, что число кластеров с двумя элементами существенно уменьшилось, соответственно, увеличилась универсальность полученных кластеров и, как следствие, быстродействие и устойчивость к случайному шуму классификации событий СОСП в целом.

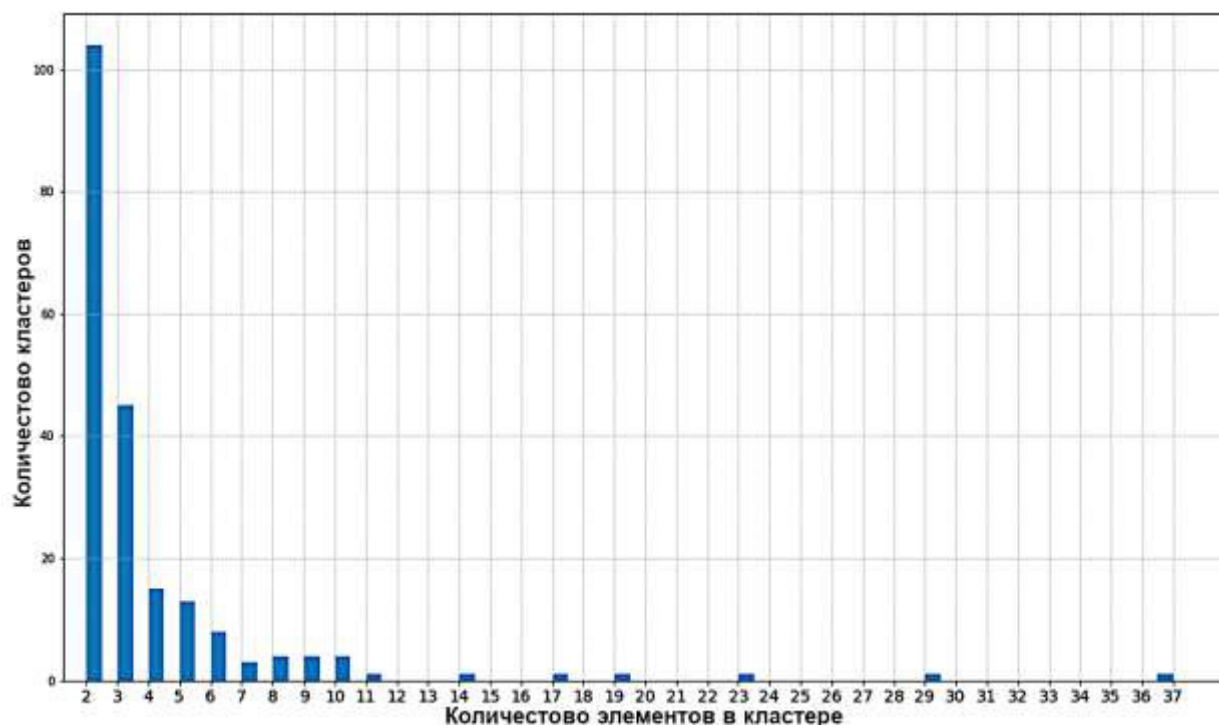


Рисунок 7 – Распределение при первичной кластеризации [Distribution in primary clustering]

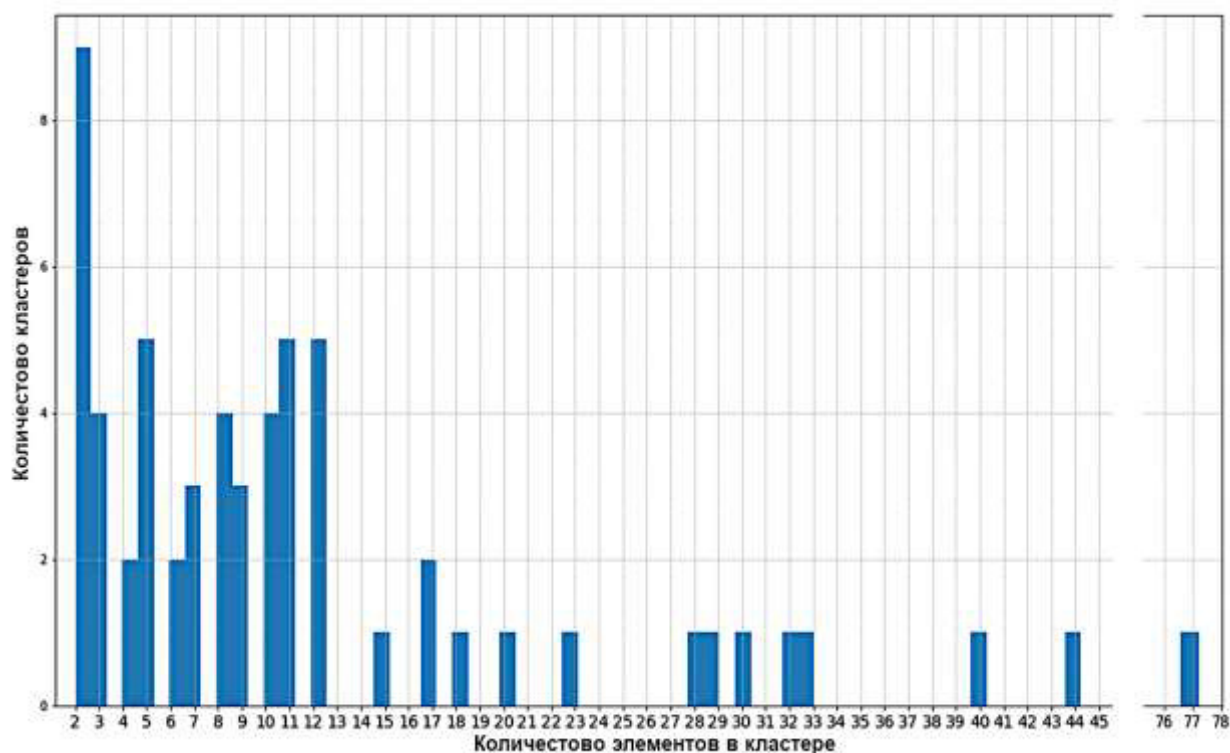


Рисунок 8 – Агломеративная кластеризация [Agglomerative clustering]

Выводы

1. Определены центры классов, соответствующих штатным операциям РУ, определены допустимые границы классов. Данные классы будут использоваться в качестве эталонов для дальнейшей классификации сигналов при работе СОСП в режиме реального времени.

2. Классификация ударов молотками произведена с высоким качеством, при весьма узких диапазонах разброса признаков. Процент классифицированных событий по БК с номерами 11, 12, 16 достаточно высок, что совокупно свидетельствует об эффективности разработанного алгоритма классификации.

3. Агломеративная кластеризация – эффективный способ борьбы с зашумлением данных и недостаточным количеством данных (заполнение пропусков в значениях признаков в предположении, что допустимые интервалы представляют собой связанные множества, это позволяет делать выводы при недостаточной статистике).

4. Разработанный алгоритм позволяет существенно уменьшить количество классов, и одновременно повышается доля кластеризованных событий, что приведет к повышению быстродействия и увеличению устойчивости к зашумлению данных СОСП в целом.

СПИСОК ИСТОЧНИКОВ

1. Центр диагностики «Диапром». Сопровождение эксплуатации. – URL: <http://www.diaprom.com/about/operation> (дата обращения: 21.11.2022).
2. Центр диагностики «Диапром». Система обнаружения свободных предметов в главном циркуляционном контуре ВВЭР-1000 (СОСП). – URL: <http://www.diaprom.com/projects/?p=2> (дата обращения: 21.11.2022).
3. Аркадов, Г.В. Системы диагностирования ВВЭР / Г.В. Аркадов, В.И. Павелко, Б.М. Финкель. – Москва : Энергоатомиздат, 2010. – 391 с.
4. Воронов, А.В. Опыт использования систем обнаружения свободных и слабозакрепленных предметов в контуре циркуляции теплоносителя реакторных установок Нововоронежской АЭС / А.В. Воронов, М.Т. Слепов // Известия вузов. Ядерная энергетика. – 2022. – № 2. – С. 15-26.
5. Максимов, И.В. Метод локализации удара для системы обнаружения свободных предметов в контуре циркуляции теплоносителя реакторных установок с ВВЭР / И.В. Максимов, В.В. Перевезенцев // Известия вузов. Ядерная энергетика. – 2019. – № 4. – С. 28-38.
6. Центр диагностики «Диапром». «Молоток импульсный ИМ-1». – URL: <http://www.diaprom.com/products/?p=14> (дата обращения: 21.11.2022).
7. Чабан, Л.Н. Теория и алгоритмы распознавания образов / Л.Н. Чабан. – Москва : Московский государственный университет геодезии и картографии, 2004. – 70 с.
8. Аркадов, Г.В. Кластеризация акустических событий в главном циркуляционном контуре реакторной установки с ВВЭР-1000/1200, обусловленных штатными технологическими операциями / Г. В. Аркадов, И. В. Трыкова, К. И. Коцоев // Глобальная ядерная безопасность. – 2022. – № 3(44). – С. 43-55.
9. Jain, A.K. Algorithms for Clustering Data / A.K. Jain, R.C. Dubes– Englewood Cliffs (NJ): Prentice-Hall, 1988. – 304 p.
10. Implementing Agglomerative Clustering using Sklearn – URL: <https://www.geeksforgeeks.org/implementing-agglomerative-clustering-using-sklearn/> (дата обращения: 21.11.2022).
11. Жамбю, М. Иерархический кластер-анализ и соответствия / М. Жамбю – Москва : Финансы и статистика, 1988. – 345 с.
12. Древовидная схема. – URL: https://ru.wikipedia.org/wiki/Древовидная_схема (дата обращения: 21.11.2022).
13. Sneath, P.H.A. Numerical taxonomy: The principles and practices of numerical classification / P.H.A. Sneath, R.R. Sokal – San-Francisco: Freeman, 1973. – 573 p. (in English).
14. Ward J.H. Hierarchical grouping to optimize an objective function / J.H. Ward // J. of the American Statistical Association. – 1963. V.58. P. 236-244.

REFERENCES

- [1] Centr diagnostiki «Diaprom» [Scientific and Technical Center «Diaprom»], Soprovozhdeniye ekspluatatsii [Operation support], <http://www.diaprom.com/about/operation> (accessed 11/21/2022) (in Russian).
- [2] Centr diagnostiki «Diaprom». [Scientific and Technical Center «Diaprom»], Sistema obnaryzheniya svobodnikh predmetov v glavnom tsyrculyatziionnom konture VVER-1000(SOSP) [System for Detecting Loose Objects in the Main Circulation Circuit of WWER-1000 (SDLO)], <http://www.diaprom.com/projects/?p=2> (accessed 11/21/2022) (in Russian).
- [3] Arkadov G.V., Pavelko V.I., Finkel B.M. Systemy diagnostirovaniya VVER [WWER Diagnostic Systems]. Moskva: Energoatomizdat [Moscow: Energoatomizdat, publishing house], 2010, 391 p. (in Russian).
- [4] Voronov A.V., Slepov M.T. Opit ispolzovaniya system obnaruzheniya svobodnikh I slabozacreplennikh predmetov v konture tsirculyacii teplonositelya reaktornikh ustanovok Novovoronezhskoy AES [Experience of Detection Systems Use for Loose Parts and Weakly Fixed Objects in the Coolant Circulation Circuit of Novovoronezh NPP Reactor Installations], Izvestiya vuzov. Yadernaya Energetika, 2022, № 2, pp. 15-26 (in Russian) DOI: 10.26583/npe.2022.2.02
- [5] Maksimov I.V., Perevezencev V.V. Metod lokalizatsii udara dlya systemi obnaruzheniya svobodnikh I slabozacreplennikh predmetov v konture tsirculyacii teplonositelya reaktornikh ustanovok s VVER [A Localization Method of Loose Part of WWER], Izvestiya vuzov. Yadernaya Energetika, 2019, № 4, pp. 28-38 (in Russian) DOI 10.26583/npe.2019.4.02
- [6] Centr diagnostiki «Diaprom» [Scientific and Technical Center «Diaprom»], Molotok impulsniy IM-1 [Impulse Hammer IM-1], <http://www.diaprom.com/products/?p=14> (accessed 11/21/2022) (in Russian).
- [7] Chaban L.N. Teoriya I algoritmi raspoznavaniya obrazov [Theory and Algorithms of Pattern Recognition]. Moskva: Moskovskiy gosudarstvenniy universitet geodezii I kartographii [Moscow: Moscow State University of Geodesy and Cartography], 2004, 70 p. (in Russian).
- [8] Arkadov G.V., Trykova I.V., Kotsoev K.I. Klasterizatsia akusticheskikh sobityi v glavnom cirkulatsionnom konture reaktornoy ustanovki s VVER-1000/1200, obuslovlennikh shtatnimi technologicheskimi operatsiyami [Clustering of Acoustic Events in the Main Circulation Circuit of a WWER 1000/1200 Reactor Facility Caused by Normal Technological Operations] // Globalnaya yadernaya bezopasnost [Global nuclear safety], 2022, № 3(44), pp. 43-55 (in Russian).
- [9] Jain A.K., Dubes R.C. Algorithms for Clustering Data – Englewood Cliffs (NJ): Prentice-Hall, 1988. 304 p. (in English).
- [10] Implementing Agglomerative Clustering using Sklearn, <https://www.geeksforgeeks.org/implementing-agglomerative-clustering-using-sklearn/> (r accessed 11/21/2022) (in English).
- [11] 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).
- [12] Dendrogram, <https://en.wikipedia.org/wiki/Dendrogram> (accessed 11/21/2022) (in Russian).
- [13] Sneath, P.H.A., Sokal R.R. Numerical Taxonomy: The Principles and Practices of Numerical Classification. San-Francisco: Freeman, 1973, 573 p. (in English).
- [14] Ward J.H. Hierarchical Grouping to Optimize an Objective Function // J. of the American Statistical Association, 1963, v.58, pp. 236-244 (in English).

**Application of Hierarchical Agglomerative Clustering to Create a Basic Classes
Database of WWER 1000/1200 Reactor Facility**

Gennadiy V. Arkadov¹, Irina V. Trykova², Denis V. Zvyagincev³, Konstantin I. Kotsoev⁴

^{1,2,3,4}Joint Stock Company "Scientific and Technical Center "Diaprom", Koroleva str., 6, Obninsk, Kaluga region, Russia 249031

⁴Bauman Moscow State Technical University, 2 Bauman str., 5/1, Moscow, Russia 105005

¹skrepka1964@gmail.com

²trykova@kvantprogramm.ru

³zvyagincev@diaprom.ru

⁴kotsoev@kvantprogramm.ru

Received by the editorial office on 11/23/2022

After revision on 01/31/2023

Accepted for publication 02/13/2023

Abstract. In reactor plants with a water-water power reactor (WWER) free, weakly fixed and foreign objects may appear in the main circulation circuit, posing a threat to the integrity of the equipment and the safety of the reactor plant. For the purpose of early detection of these objects, the NPP is equipped with a system for detecting loose/weakly fixed objects (SDLO). In addition to the detection of loose/weakly fixed objects, the functions of the SDLO include the classification of registered events. The possibility of applying the classification algorithm is based on the fact that the signals from the operation of standard equipment are highly repeatable, even in the presence of noise, while a free object is characterized by a large stochastic component and its own deterministic class cannot be formed for it. Classification reduces the number of false alarms, allowing you to select signals from regular operations, while signals from one process must be assigned to one class. The idea of the article is to "train" SDLO on a certain archive of data characterizing the normal functioning of the reactor plant, create a library of "base" classes and set the boundaries of each class so that, on the one hand, take into account the possible variability of signal parameters due to noise. Having defined the base classes, we can state that if a newly received signal falls into one of the classes, then it reflects the regular operation of the RI, while signals that do not fall into any of the classes may be the result of the appearance of a free/weakly fixed object. The article analyzes a lot of events accumulated in the archive of one of the existing SDLO. Their clustering was carried out, as a result of which the classes of events corresponding to regular technological operations were identified. For each class, the center of the class and the allowable limits of deviations from the center are calculated. All class centers obtained are benchmarks against which the SDLO either classifies a newly detected event in real time or characterizes it as "unclassified".

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

For citation: Arkadov G.V., Trykova I.V., Zvyagincev D.V., Kotsoev K.I. Application of Hierarchical Agglomerative Clustering to Create a Basic Classes Database of WWER 1000/1200 reactor facility // Global nuclear safety. 2023. No. 1(46). P. 54-66 <http://dx.doi.org/10.26583/gns-2023-01-05>.