

ОНТОЛОГИЧЕСКИЕ МОДЕЛИ ТРАНСФОРМАЦИИ ДАННЫХ О СОСТОЯНИИ ТЕХНИЧЕСКИХ ОБЪЕКТОВ

Н.А. Жукова

*Санкт-Петербургский институт информатики и автоматизации РАН, Санкт-Петербург, Россия
nazhukova@mail.ru*

Аннотация

Для наблюдения за техническими объектами на них устанавливается множество датчиков, которые позволяют собирать данные о состоянии этих объектов. В зависимости от свойств собираемых данных и решаемых практических задач определяются процессы их обработки. При изменении данных или требований к результатам их обработки процессы перестраиваются. Для обеспечения эффективной обработки предложены метод и модель трансформации данных. Они предусматривают формальное описание последовательностей трансформаций в общем виде, а также их представление в виде процессов обработки. Для практической реализации метода предложено представлять трансформации в виде связанных онтологических моделей, включающих функциональную, информационную и процессно-сервисные модели. Построение моделей осуществляется с использованием правил, множество которых представляется в виде управляющей модели. Приводится онтологическая модель для трансформации данных, представленных в виде временных рядов. Структуру её классов определяют классификаторы, определённые для исходных временных рядов и их представлений, а также методов, алгоритмов и процедур их обработки и оценки результатов. Приводится пример обработки результатов измерений значений параметров давления технического объекта космического назначения при проведении технического контроля его состояния.

Ключевые слова: модели трансформации, обработка данных, технические объекты, временные ряды.

Цитирование: Жукова, Н.А. Онтологические модели трансформации данных о состоянии технических объектов // Онтология проектирования. – 2019. – Т. 9, №3(33). - С. 345-360. – DOI: 10.18287/2223-9537-2019-9-3-345-360.

Введение

От технических объектов поступают потоки данных, содержащие разнородные результаты измерений значений параметров этих объектов. Сырые данные трудно применимы при прогнозировании состояния объектов, управлении ими, а также решении других прикладных задач. Использование данных становится возможным после некоторой предварительной обработки - трансформации для представления в нужном виде. Методы трансформации данных отличаются показателями эффективности. В качестве таких показателей используются погрешность, время трансформации, требуемый объём памяти и другие. При выборе методов трансформации должны учитываться условия обработки данных, включая наличие соответствующих программ, технических средств и т.п. Таким образом, имеет место актуальная научная задача эффективной трансформации данных, получаемых от наблюдаемых объектов, к требуемому виду. Эта задача относится к области интеллектуальной обработки и анализа данных. Несмотря на достигнутые успехи, многие аспекты автоматической трансформации данных к требуемому виду оставлены без внимания. Значительная часть предложенных решений носит частный характер.

В работе предлагаются новые многоуровневые модели и методы трансформации данных, позволяющие выполнять управляемое построение и перестройку процессов обработки данных в различных условиях. Предлагаемые модели и методы предусматривают построение различных представлений данных, а также многократное преобразование построенных представлений. При этом используются методы статистической обработки данных, интеллектуального анализа, машинного обучения. Способы представления данных об объектах определяются требованиями, предъявляемыми к результатам обработки, и свойствами исходных данных. Выбор алгоритмов обработки осуществляется исходя из формируемых представлений данных.

В состав разработанных моделей и методов входят метод и модель многоуровневой трансформации данных, а также метод и модель обеспечения многоуровневой трансформации. Метод трансформации позволяет строить многоуровневые модели, содержащие формальное описание последовательностей трансформаций и их представлений. Для построения моделей трансформации разработан метод обеспечения трансформации данных. Он предусматривает построение модели обеспечения, которая описывает формальные модели в виде системы функциональных, информационных, процессных и сервисных моделей, и управляющей модели, которые вместе представляются в виде онтологических моделей.

1 Метод многоуровневой трансформации данных

Предлагаемый метод многоуровневой трансформации данных предусматривает, что известны исходные данные, условия их обработки, а также требования, предъявляемые к результатам обработки.

Метод включает выполнение четырёх последовательных шагов.

- 1) определение пространства построения процессов обработки, описание процесса обработки в общем виде. При этом определяются стратегии для достижения цели обработки, а также следующие пространства: фазовое; управления обработкой; функционалов, применимых для достижения цели обработки.
- 2) построение семейств трансформаций данных для выбранной стратегии. Строятся множества допустимых правил управления трансформациями для выбранной стратегии, формируется общая структура системы трансформаций в виде последовательности управляемых переходов между различными представлениями данных.
- 3) построение системы трансформаций данных. Для этого определяются трансформации для типов представлений данных. На их основе строятся многоуровневые горизонтально-вертикальные структуры трансформаций.
- 4) построение параметрических семейств процессов обработки данных. На основе построенных трансформаций с горизонтально-вертикальной структурой определяются семейства параметрических процессов обработки данных.

2 Многоуровневая модель трансформации данных

Исходное пространство трансформаций данных определяется допустимым контентом информационных потоков, допустимым контекстом обработки, множеством функционалов, применимых для обработки контента в контексте, возможными стратегиями обработки. Под стратегией понимается совокупность правил выбора действий. Процессы, реализующие некоторую стратегию обработки за конечное время, представляются в виде множества элементарных процессов обработки, для которых определены правила управления ими.

Для построения процессов обработки определяются параметрические семейства трансформаций. Семейства трансформаций обеспечивают достижение конечной цели за m итераций за счёт управляемых переходов на множестве допустимых правил управления для заданной стратегии:

$$F_k = T_{ck} \otimes F_{k-1}, \quad k = \overline{1, m},$$

где F_k - правила управления на k -ом шаге; T_{ck} - состав трансформаций на k -ом шаге, \otimes - операция трансформации. Такие переходы можно рассматривать как управляемое блуждание на множестве допустимых правил управления. При переходах могут строиться новые представления обрабатываемых данных или преобразовываться имеющиеся. В первом случае трансформации, предусматриваемые переходом, относятся к горизонтальным, во втором - к вертикальным.

В результате семейства трансформаций представляют собой структуры вида:

$$T = \{T^G \cup T^V\},$$

где T^G - вертикальные, T^V - горизонтальные трансформации. Эти структуры определяют структуры формируемых моделей трансформации данных.

На основе семейств трансформаций строятся системы трансформаций. При этом построенные семейства трансформаций уточняются, исходя из представлений обрабатываемых данных, между которыми осуществляются переходы. Системы трансформаций составляют основу для построения параметрических семейств вертикальных и горизонтальных процессов обработки. При определении семейств процессов обработки учитывается контент и контекст обработки.

Предложенная модель позволяет описывать процессы обработки данных в общем виде и решать различные задачи обработки данных, в том числе задачи слияния данных [1-3].

3 Обеспечение многоуровневой трансформации данных

3.1 Метод обеспечения многоуровневой трансформации данных

Метод предусматривает построение многоуровневых представлений процессов обработки. Уровни представлений процессов обработки соответствуют уровням, определённым в модели трансформации.

Предлагаемый метод предусматривает построение следующих представлений процессов обработки данных [4].

- 1) на верхнем уровне процессы обработки описываются в виде иерархии этапов и подэтапов обработки. Этапы и подэтапы определяются исходя из решаемой задачи и входящих в её состав подзадач.
- 2) на следующем уровне этапы обработки представляются в виде технологий обработки. Технологии представляют собой обобщённые методы решения. Выбор технологий осуществляется с применением функций, позволяющих выбирать технологии обработки из множества возможных с использованием правил.
- 3) на третьем уровне определяются шаблоны обработки. При этом каждая технология представляется в виде множества связанных шаблонов. Под шаблоном обработки понимается детальное описание реализации технологии, представляемое в виде графа. В его узлах могут размещаться шаблоны или группы алгоритмов обработки. Для формального описания шаблонов применяются ориентированные атрибутированные графы отношений. Атрибуты являются типизированными. С каждым атрибутом

ассоциирован вектор характеристик. Выделяются два типа атрибутов вершин: атрибуты, которые содержат именованные группы алгоритмов обработки, и атрибуты, содержащие именованные шаблоны обработки. Атрибутами рёбер являются правила переходов между вершинами. Для задания атрибутов применяются функции назначения атрибутов вершинам и рёбрам.

- 4) на четвёртом уровне моделей определяются процессы обработки, представляемые последовательностями алгоритмов обработки. При этом из каждой группы алгоритмов, определённой на третьем уровне, выбирается один или несколько алгоритмов. Алгоритмы имеют описания, которые включают базовые сведения об алгоритмах, описание условий применения и критерии оценки, *пред-* и *пост-* условия выполнения. Определяются правила и условия выполнения алгоритмов. Алгоритмы могут выполняться последовательно или параллельно.

В соответствии с процессами обработки определяются исполняемые процессы обработки данных, для которых определены все параметры их выполнения.

3.2 Модель обеспечения многоуровневой трансформации данных

Модель обеспечения имеет многоуровневую структуру. Уровни модели соответствуют различным представлениям трансформаций.

Каждое представление рассматривается с нескольких точек зрения. Основными из них являются функциональная, информационная и процессно-сервисная. Другие возможные точки зрения рассмотрены в [5]. Для всех точек зрения строятся соответствующие им модели.

Функциональные модели описывают задачи, решаемые при обработке, состав необходимых для этого входных данных и ожидаемые результаты обработки. Информационные модели содержат исходные данные и описания результатов их обработки. Процессно-сервисные модели определяют состав преобразований, которые необходимо выполнить для решения поставленных задач обработки, и программные сервисы, которые позволяют их реализовывать. Функциональные, информационные и процессно-сервисные модели связаны между собой. Модели функциональных, информационных и процессно-сервисных представлений трансформаций и установленные между ними взаимосвязи показаны на рисунке 1.

На нижнем уровне элементами информационных моделей являются элементы исходных данных, полученных от объектов. Процессные модели представлены в виде исполняемых процессов, определены сервисы, позволяющие их выполнять. Функциональные модели описывают решаемые задачи в виде множества отдельных элементарных функций обработки.

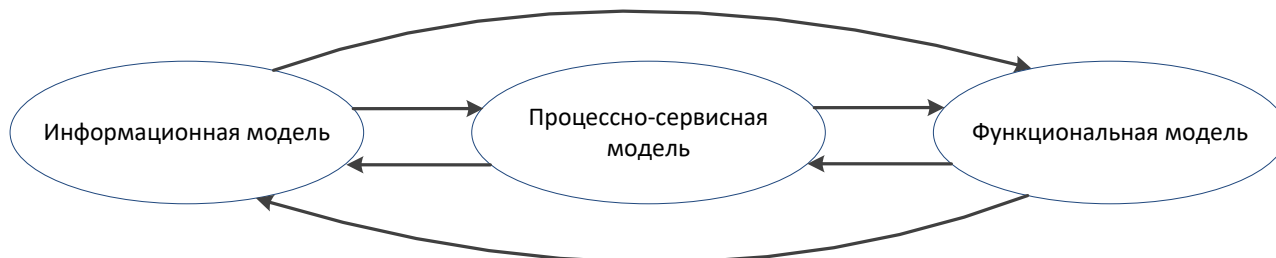


Рисунок 1 – Модели представления трансформаций данных

Правила построения моделей определяются в управляющей модели. Правилами управляющей модели определяются возможные переходы между функциональными, информационными и процессно-сервисными моделями. Связи являются двунаправленными и позволяют:

- по данным, содержащимся в информационных моделях, определять задачи обработки, которые могут решаться, т.е. определять элементы функциональных моделей;
- для решаемых задач обработки и с учётом имеющиеся исходных данных определять возможные процессы обработки данных;
- определять новые информационные модели или детализировать существующие за счёт новых информационных элементов, получаемых в результате обработки исходных данных.

Обратные связи определяют накладываемые на модели ограничения. Информационные и процессно-сервисные модели строятся с учётом ограничений, которые определяются решаемыми задачами обработки. Ограничения формулируются исходя из требований, которые предъявляются пользователями к результатам обработки данных. Информационные модели могут ограничивать возможности построения процессов обработки. При построении процессов обработки некоторые алгоритмы обработки могут оказаться неприменимы из-за нехватки исходных данных.

3.3 Онтологическая модель многоуровневой трансформации данных

Онтологические модели строятся в соответствии со структурой и взаимосвязями, которые определены для модели обеспечения. В условиях, когда требуется решение ограниченного числа задач, онтологические модели строятся для фрагментов моделей обеспечения.

Структура онтологической модели приведена на рисунке 2.

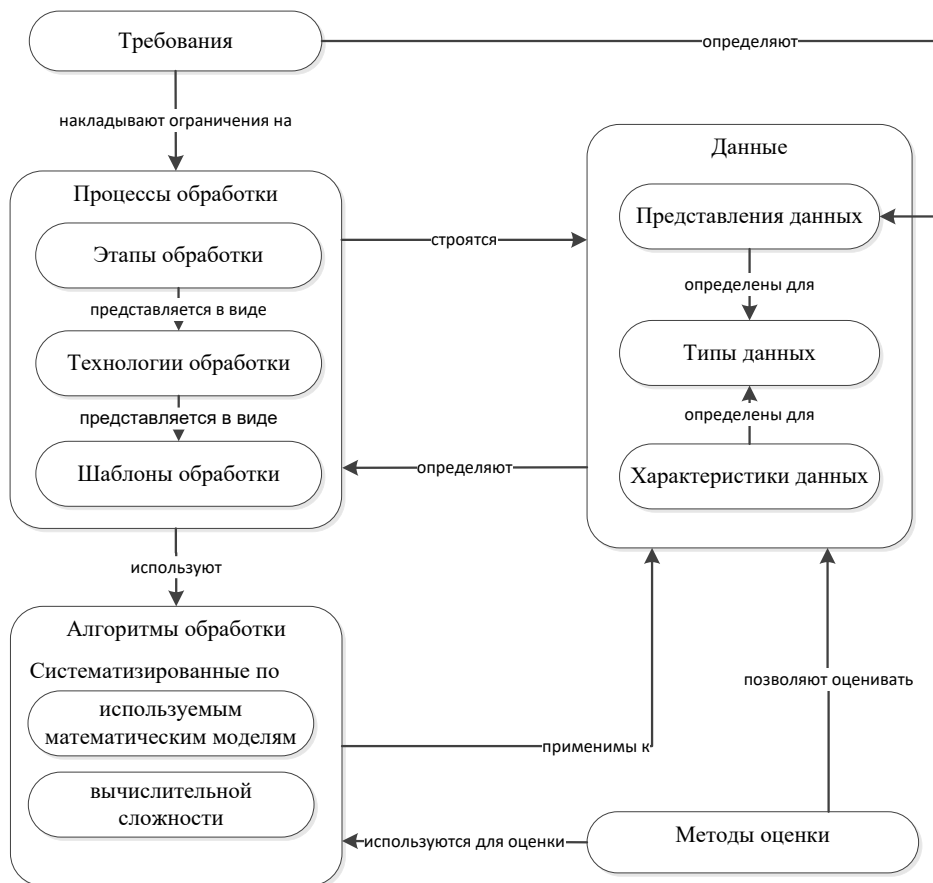


Рисунок 2 – Структура онтологической модели трансформации данных

В приведённой на рисунке 2 структуре процессно-сервисная модель описывается тремя классами – «Процессы обработки», «Алгоритмы обработки», «Методы обработки», функциональная модель представлена классом «Требования», а класс «Данные» определён в соответствии с информационной моделью. Правила управляющей модели отражены в виде связей, установленных между классами.

Классы онтологической модели имеют множество подклассов. Структуры классов определяются в соответствии с известными классификаторами тех объектов, которые к этим классам отнесены. При построении классов могут использоваться один или несколько классификаторов. Классы, построенные на основе одного классификатора, имеют иерархическую структуру. При учёте различных классификаторов структуру классов определяют несколько иерархий. При этом объекты принадлежат классам из различных иерархий.

Класс «Требования» определяет требования, которые могут предъявляться к результатам обработки данных со стороны потребителей. Они определяются исходя из задач, решаемых конечными пользователями. В требованиях также учитываются ограничения, накладываемые на формируемые процессы обработки данных.

Класс «Процессы обработки» содержит классы, определяющие элементы, используемые для описания различных представлений процессов, в частности, их представления в виде этапов обработки, технологий и шаблонов обработки, а также в виде последовательностей алгоритмов обработки.

Класс «Данные» содержит информацию о возможных типах данных, их свойствах, а также о представлениях данных, которые могут быть получены в результате их обработки.

Класс «Алгоритмы обработки» содержит описания известных алгоритмов обработки данных. В модели алгоритмы классифицируются по нескольким основаниям, в частности, по используемым математическим моделям и по вычислительной сложности. Возможны классификации по другим основаниям.

Класс «Методы оценки» содержит описание моделей и методов, которые могут применяться для оценки результатов обработки данных с применением алгоритмов из класса «Алгоритм обработки».

В модели определены свойства данных и объектные свойства. Свойства данных применяются для описания значений отдельных параметров процессов, данных, алгоритмов. Например, для алгоритмов определяются их параметры, диапазоны значений параметров, задаются *пред-* и *пост-* условия, размещаются неформальные описания алгоритмов, описываются программные сервисы, реализующие эти алгоритмы и т.д.

К объектным отнесены свойства, которые связывают классы между собой и используются при построении процессов обработки. В модели определены следующие объектные свойства.

- Свойства «Накладывают ограничения на» и «Определяют» связывают класс «Требования» с классами «Процессы обработки» и «Данные». За счёт этих свойств задаются ограничения, накладываемые на процессы обработки, и определяются представления данных, которые требуется получить в результате обработки данных.
- Свойство «Определены для» отражают связи между типами данных и их характеристиками, а также между типами данных и их возможными представлениями. За счёт этого свойства определяются возможные способы описания данных, которые могут быть получены в результате их обработки.
- Свойства «Определяют» и «Строятся» устанавливают связи между классами «Процессы обработки» и «Данные». Определение этапов обработки осуществляется исходя из требуемых представлений данных. Выбор технологий и шаблонов основан на оценке характеристик обрабатываемых данных.
- Свойство «Представляются в виде» обеспечивает возможность представления этапов процессов обработки в виде технологий обработки и шаблонов обработки.
- Свойства «Используют» и «Применимы к» позволяют определять алгоритмы, используемые в процессах обработки данных. Свойство «Применимы к» позволяет выбирать алгоритмы с учётом характеристик обрабатываемых данных.
- Свойства «Позволяют оценивать» и «Используются для оценки» дают возможность оценивать исходные данные и результаты их обработки. При оценке результатов обработки учитываются алгоритмы, с применением которых выполнялась обработка.

Построение процессов обработки на основе онтологической модели предполагает выполнение следующих шагов.

- 1) формализация требований пользователей, условий обработки данных. Определение на их основе требуемого представления данных, которое необходимо построить, назначение ограничений для параметров формируемых процессов обработки.
- 2) определение типов исходных данных и рассчитываемых для них характеристик. Определение промежуточных представлений данных, позволяющих перейти от исходных данных к требуемому представлению.
- 3) определение этапов обработки, необходимых для формирования каждого из представлений данных. Выбор технологий обработки, шаблонов обработки и алгоритмов обработки. В случае, если алгоритмы из класса «Алгоритм» не позволяют обеспечить требуемую обработку, то возможно два альтернативных варианта. В первом варианте исключается из рассмотрения одно или несколько ограничений, накладываемых на

параметры процессов обработки. Второй вариант предполагает пересмотр требований к процессам. С использованием информации о программных сервисах, обеспечивающих реализацию алгоритмов, процесс описывается в виде последовательности вызовов сервисов.

- 4) полученные результаты обработки оцениваются с использованием методов и моделей из класса «Методы оценки». В случае, если результаты обработки не отвечают требованиям, то процессы обработки перестраиваются.

4 Онтологическая модель обработки временных рядов

Онтологическая модель обработки временных рядов результатов измерений значений параметров технических объектов строится на основе классификаторов, определённых для исходных временных рядов и их представлений, а также методов, алгоритмов и процедур их обработки и оценки результатов. Классификатор требований в рамках онтологической модели не рассматривается. Этот классификатор определяется исходя из решаемых прикладных задач и не зависит от типа обрабатываемых данных.

Классификатор временных рядов измерений. Классификация временных рядов измерений основана на формировании групп временных рядов в соответствии с их поведением. В зависимости от характера изменения во времени параметры разделяются на функциональные, сигнальные и константные [6]. Характерная особенность функциональных параметров состоит в том, что они являются непрерывными во времени функциями. Для сигнальных параметров характерно скачкообразное изменение во времени, например, связанное с переходом из одного дискретного состояния в другое. Константные параметры описываются одним значением. В зависимости от скорости изменения во времени функциональные параметры разделяются на медленно меняющиеся и быстро меняющиеся [6].

По характеру поведения временных рядов результатов измерений выделяются стационарные, нестационарные и кусочно-стационарные временные ряды. Временные ряды измерений независимо от их физической природы в большинстве случаев являются случайными нестационарными процессами. Однако для них могут быть определены относительно протяжённые участки стационарности.

Для временных рядов измерений медленно меняющихся функциональных параметров определяется наличие или отсутствие разрывов в первой и второй производных.

Для функциональных параметров определяется возможность описания результатов их измерений, представленных в виде временных рядов, с применением параметрических моделей. Для нестационарных рядов подбор модели осуществляется для каждого из стационарных сегментов ряда. Состав моделей определяется в соответствии с существующими статистическими и интеллектуальными моделями описания временных рядов. Для оценки соответствия модели и временного ряда в типовом случае используется метод наименьших квадратов, возможно использование любого другого метода, например, метода максимального правдоподобия.

Наиболее распространенным является способ классификации временных рядов, основанный на вычислении совокупности различных характеристик исходного временного ряда. Состав рассчитываемых характеристик определяется соответствующим классификатором.

Классификатор сегментов временных рядов. Сегментирование осуществляется для кусочно-стационарных и нестационарных временных рядов. Рассматривается два основных способа определения сегментов.

Первый способ предполагает анализ диапазона значений одной или совокупности рассчитанных для сегмента характеристик. Например, временные ряды значений константных параметров могут описываться медианой, разбросом значений, а также процентным соотношением значений, совпадающих со значением медианы, и отличных от неё. При таком способе сегментации формируемые сегменты классифицируются по рассчитываемым характеристикам.

В соответствии со вторым способом определения сегментов осуществляется поиск схожих сегментов в априорно сформированной базе типовых сегментов. В базу типовых сегментов могут включаться сегменты, соответствующие константному, линейно возрастающему / убывающему, выпукло / вогнуто возрастающему / убывающему поведению временного ряда. В процессе накопления исторических данных база сегментов расширяется за счет добавления в базу специализированных сегментов, отражающих характерные особенности поведения временных рядов, полученных в результате измерений значений параметров объектов.

При поиске сегментов целесообразным является использование мульти-разрешающего подхода к представлению временных рядов. В соответствии с ним верхний уровень представления отражает общие тенденции в поведении временного ряда. На более низких уровнях учитываются особенности его поведения. При использовании мульти-разрешающего подхода одним из классификационных признаков сегментов является уровень представления временных рядов.

Классификатор характеристик временных рядов. Характеристики предназначены для описания временных рядов, а также их сегментов. Классификатор характеристик временных рядов предполагает формирование групп характеристик по скорости их вычисления и области представления временных рядов результатов измерений (временная, частотная, частотно-временная), а также степени информативности характеристик.

К первой группе характеристик относятся: частота измерений, значения статистических характеристик (медиана, мода, размах, ранг, среднее квадратическое отклонение, коэффициент вариации, моменты (математическое ожидание, дисперсия, асимметрия, эксцесс), поведение кривой, соответствующей временному ряду во временной области (вариабельность, ошибка кусочно-постоянной аппроксимации, ошибка кусочно-линейной аппроксимации, ошибка аппроксимации полиномом второй и более степеней, значения характерных точек, кривизна), поведение кривой в частотной области (вектор коэффициентов Фурье), поведение кривой в частотно-временной области (вектор вейвлет-коэффициентов), энтропия, вариативность производной и др.

Вторую группу определяют характеристики, которые рассматривают временной ряд как случайный стационарный процесс: одномерная и многомерная функции распределения, одномерная и многомерная плотность вероятности случайного процесса, распределение вероятностей случайных дискретных величин, спектральная плотность.

Расширенное признаковое пространство описания исходных и преобразованных временных рядов результатов измерений во временной области предусматривает вычисление производных с применением метода конечных разностей, нижней и верхней огибающих временных рядов и вариаций построенных огибающих, длины кривой, а также преобразование временного ряда с применением метода главных компонент [7]. Для преобразованных данных рассчитываются ошибки их описания константными значениями, линейными функциями, квадратичными функциями. Возможна интерполяция временных рядов кубическими сплайнами или аппроксимация заданными функциями (степенной, показательной, логарифмической, пользовательской). В этом случае в качестве характеристик рассматриваются ошибки интерполяции и аппроксимации. Характеристиками

временных рядов также являются локальная, глобальная и взвешенная сложность, показатели вариабельности, максимальное и минимальное значение кривизны и её медиана, площадь фигуры, ограниченной кривой и линией, соединяющей её крайние точки [8].

В качестве альтернативного подхода к построению классификатора характеристик может использоваться подход, основанный на типе рассчитываемых характеристик. Выделяются три основных типа характеристик – линейные характеристики, нелинейные характеристики и другие характеристики, формируемые на основе вычисления различных мер. В составе каждой из этих групп выделяется несколько подгрупп.

При определении характеристик, относящихся к линейным, используются меры, основу которых составляет вычисление линейной корреляции, частоты, построение авторегрессионных моделей. В первую подгруппу линейных характеристик входят меры, основанные на вычислении линейной и монотонной автокорреляционных функций. В зависимости от видов шкал могут вычисляться автокорреляционные функции Пирсона, определяться ранговые корреляции Кендалла или Спирмана. При вычислении мер часто рассматриваются различные временные задержки. К первой подгруппе относятся также характеристики, основанные на вычислении куммулятивной автокорреляции и частичной автокорреляции [9]. Подгруппа мер, вычисляемая при частотном представлении временных рядов, включает распределение энергии в частотных интервалах, медиану частоты, вычисленную на основе оценки спектральной мощности с применением стандартной периодограммы [10]. К третьей подгруппе относятся меры, формируемые на основе построения линейных моделей временных рядов: авторегрессионной модели, авторегрессионной модели скользящего среднего [11]. При вычислении мер строятся модели различных порядков. На основе построенных моделей могут быть вычислены следующие меры: абсолютный средний процент ошибки, нормализованное среднеквадратичное отклонение ошибки, нормализованный средний квадрат ошибки, коэффициент корреляции.

К основным группам мер, применяемым для вычисления нелинейных характеристик, относятся меры, основанные на вычислении нелинейной корреляции, размерности и сложности временного ряда, а также построении нелинейных моделей временных рядов. К первой подгруппе нелинейных характеристик относятся две меры – бикорреляционная мера (трёхточечная автокорреляция) и мера взаимной информации. Для вычисления взаимной информации применяется разбиение на равные интервалы по расстоянию и по вероятности распределения значений [12, 13]. Так же как и для корреляционных мер, вычисляются меры куммулятивной бикорреляции и куммулятивной взаимной информации для различных временных задержек [14, 15]. К подгруппе мер, основанных на вычислении размерности и сложности, относятся: размерность корреляции [16], куммулятивная плотность и обратная куммулятивная плотность. Мерами, оценивающими сложность, являются меры аппроксимации энтропии и алгоритмической сложности. При вычислении мер, основанных на нелинейных моделях, рассматриваются следующие типы моделей: локальная модель среднего; локальная линейная модель, параметры которой оцениваются с использованием метода наименьших квадратов; локальная линейная модель, параметры которой оцениваются на основе расчета регрессионной функции для главных компонент. Для оценки моделей используются характеристики, аналогичные линейным моделям.

Меры, на основе которых определяется третья группа характеристик, подразделяются на две подгруппы: статистические меры и статистические меры для осциллирующих характеристик. К первой подгруппе относятся: среднее, медиана, стандартное отклонение, гладкость, эксцесс, первый и второй моменты временного ряда, параметры Херста и корреляционные меры, которые рассчитываются с использованием метода анализа диапазонов с изменением масштаба (*Rescaled range analysis*) и метода анализ колебаний

(*Detrended actuation analysis*). Вторая подгруппа включает меры для описания осциллирующих временных рядов, в частности, могут рассчитываться локальный минимум, локальный максимум, размах, продолжительность осцилляции. Другие меры, предложенные, например, в [17].

Классификатор вариантов представления временных рядов. Варианты формализованного описания определяются для стационарных, кусочно-стационарных и нестационарных временных рядов. Рассматриваются два основных способа представления временных рядов – представления, адаптируемые к описываемым данным, и представления, не адаптируемые к данным.

При использовании адаптируемых представлений временные ряды могут описываться следующими характеристиками: коэффициенты кусочно-линейной аппроксимации и адаптивной кусочно-константной аппроксимации; коэффициенты сингулярного разложения; символьные представления временных рядов.

При использовании не адаптируемых представлений вычисляется следующий набор характеристик: коэффициенты вейвлет-преобразования с использованием ортонормированных и биортонормированных преобразований; коэффициенты спектрального представления временных рядов, полученные с использованием дискретного Фурье-преобразования и дискретного косинусного преобразования; результаты кусочно-совокупного приближения временного ряда и другие.

Классификатор методов, алгоритмов и процедур обработки и анализа данных может строиться исходя из применяемого математического аппарата. Основу такого классификатора составляют стандартные классификаторы методов цифровой обработки сигналов, математической статистики и методов искусственного интеллекта, включая методы интеллектуального анализа данных и машинного обучения.

Другой классификатор методов, алгоритмов и процедур может быть построен исходя из основных этапов обработки. Типовые процессы обработки данных и их этапы рассмотрены в [18]. Процесс обработки может быть ориентирован на получение дополнительной информации об обрабатываемых результатах измерений. В этом случае широко применяются алгоритмы разведочного анализа данных. Это могут быть наборы тестов, включая тесты оценки качества измерений, определения типов временных рядов результатов измерений, определения набора характеристик, которыми они обладают и т.д. Основу тестов составляют различные статистические процедуры. При комплексной обработке применяются методы интеллектуального анализа, в частности, сегментации, кластерного анализа, классификации, секвенциального анализа, поиска ассоциативных правил и другие. Комплексная обработка предполагает определение основных характеристик обрабатываемых временных рядов на основе выявления схожих по поведению временных рядов с известными характеристиками.

Классификатор критериев оценки данных, а также результатов работы алгоритмов. Рассматриваются три группы критериев оценки данных:

- формируемые исходя из сведений об источнике данных;
- получаемые в результате сравнения данных с историческими данными;
- рассчитываемые с применением специализированных процедур.

Состав критериев, относящихся к первой группе, представляет собой фиксированный набор рассчитываемых характеристик и допусков на их значения. Вторая группа критериев основана на определении степени схожести обрабатываемых данных и исторических данных, полученных в аналогичных условиях. Третья группа включает стандартные критерии оценки, определённые для алгоритмов обработки данных. Например, оценки результатов кластерного анализа могут основываться на вычислении критериев

компактности данных внутри кластера и удаленности между элементами различных кластеров. Для этого рассчитываются индекс Данна и подобные ему, индекс Девиса-Болдиена, среднеквадратичное отклонение и коэффициент детерминации, SD индекс, S_Dbw индекс. Для других групп алгоритмов имеются свои критерии. Для алгоритмов ассоциации критерии рассмотрены в [19].

Пример онтологической модели для обработки временных рядов рассмотрен в [20].

Рассмотренная онтологическая модель может быть расширена за счёт других существующих, а также вновь разрабатываемых моделей и методов. Для расширения онтологической модели могут использоваться открытые информационные ресурсы, в частности, [21].

5 Пример применения онтологической модели при обработке данных технических объектов

В области объектов космической техники перед проведением пусков предусматривается проведение комплексных проверок их состояния. В ходе проверок инженерами-анализаторами проводится детальное исследование поведения всех систем объектов при различных условиях. В ходе проверок проводятся измерения значений параметров объектов. Задача состоит в выявлении по результатам измерений параметров таких агрегатов и узлов, реакция которых на внешние воздействия не в полной мере соответствует ситуации. Отказ элементов технических изделий в основном наблюдается при переходе из одного состояния в другое. Таким образом, наибольший интерес с точки зрения анализа состояния объектов представляют переходные процессы. При этом, именно эти процессы, являются наиболее сложными для обработки.

На рисунке 3 приведены результаты измерений параметров давлений продувок магистралей горючего (ПМГ) ракетносителя «Союз-2» для заправочных магистралей четырёх блоков («а»-«г»). Блоки имеют одинаковую конструкцию, соответственно, поведение одинаковых параметров на различных блоках должно быть схожим. Просмотр графиков изменения давлений не даёт представления о различиях в поведении параметров. Для выявления расхождений требуется точное сопоставление данных и их детальный совместный анализ.

На рисунке 4а показаны результаты измерений параметров ПМГ на одном графике. На рисунке 4б представлены фрагменты графиков, на которых наблюдаются расхождения, в увеличенном масштабе. Процессы обработки основываются на последовательном применении процедур сегментации и кластерного анализа. В результате сегментации временные ряды результатов измерений представляются в виде последовательности стационарных фрагментов. Кластерный анализ позволяет выявить отличия в поведении параметров за счёт сопоставления фрагментов различных параметров между собой. Подробное описание исходных данных и процессов их обработки приведено в [22].

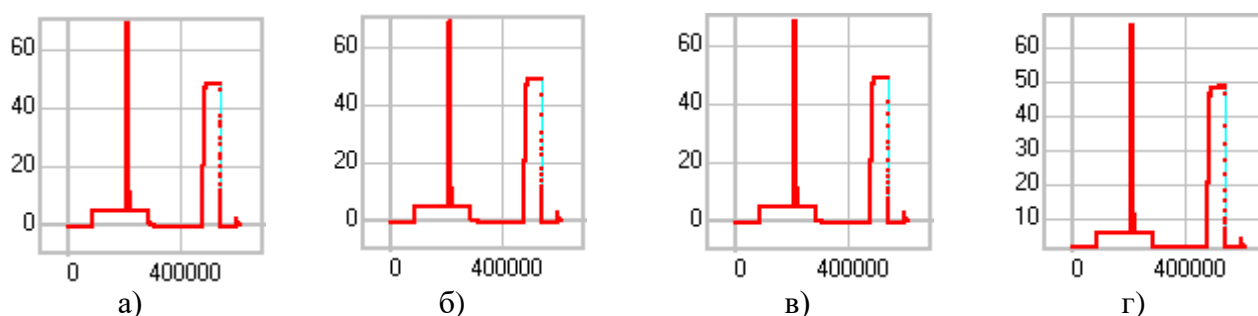


Рисунок 3 – Результаты измерений параметров давлений ПМГ для заправочных магистралей четырёх блоков

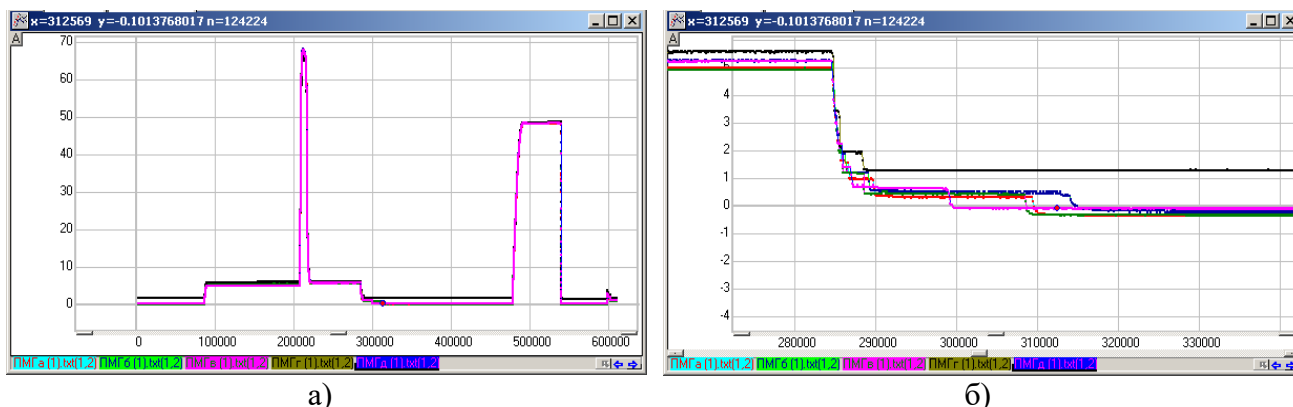


Рисунок 4 – Расхождение в поведении параметров ПМГа- ПМГг

При использовании предложенной модели и метода трансформации данных процесс обработки строится следующим образом.

Сегментация выполняется за несколько шагов. Преобразования на каждом шаге выбираются исходя из свойств исходных данных с использованием классификаторов. В соответствии с классификатором временных рядов результаты измерений параметров давления относятся к функциональным медленно меняющимся параметрам. По характеру поведения временные ряды результатов измерений значений параметров являются кусочно-стационарными. Согласно классификатору сегментов временных рядов они могут представляться в виде последовательности типовых сегментов. При сегментации выполняется аппроксимация фрагментов исходных временных рядов полиномами первой и второй степеней. Выбор метода осуществляется на основе классификатора методов, алгоритмов и процедур обработки и анализа данных. Выделенные сегменты в соответствии с классификатором характеристик временных рядов описываются коэффициентами аппроксимирующих полиномов. Для выбора методов оценки результатов сегментации используется классификатор критериев оценки результатов работы алгоритмов. При кусочно-полиномиальной аппроксимации временных рядов результаты могут оцениваться с помощью метода наименьших квадратов.

Выбор алгоритмов кластерного анализа также осуществляется с использованием классификатора методов, алгоритмов и процедур обработки и анализа данных. Кластеризация векторов полиномиальных коэффициентов осуществляется на основе расчёта расстояний между векторами. К этой группе алгоритмов кластеризации относится алгоритм К-средних. При кластеризации векторов, имеющих большую размерность, результаты кластеризации представляются в пространстве главных компонент.

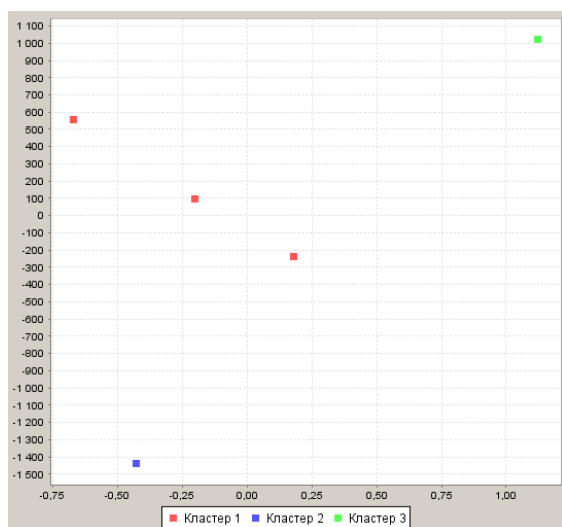
В результате сегментации в рассматриваемых параметров ПМГ выделяется шесть сегментов. На рисунке 5 показаны коэффициенты, рассчитанные для сегментов параметра ПМГа. Аналогичные коэффициенты рассчитываются для сегментов параметров ПМГб - ПМГг.

'05-11-04','009.I','ПМГа',0, -0.00000000000017, 0.00000001070337, -0.39169607280255,0, -0.000093428, 8.177785383, -238598.966329420,0, -0.00000000001629, 0.00000253145799, 4.89322082561388,0, -0.0025294, 530.3601456, -37067947.1879784,0, -0.000025531, 5.446707953, -387210.622801123,0, 0.0008606, -188.4030999, 13748134.1363760,0, 0.00000000001283, -0.00000313377703, 5.33150154998261,0, 0.00000013424, -0.04151415427, 4279.45720458188,0, 0.00000000000046, -0.00000020657920, -0.36096701511691,0, 0.00001342, -6.38363600, 1011109.54211471,0, -0.00000000001507, 0.00000766481145, 47.05548231648211,0,

81.96, -44283782.97, 7975639317406.97,0, 0.0000000001306, -0.00000742067596, 1.01376832581661,0, -2.304,
1382177.784, -276408065663.900,0, 0.00002792,
-16.83409530, 3383403.13271503,0, -0.000000004467, 0.002721174969, -552.265783450761

Рисунок 5 – Описание сегментов параметра ПМГа в виде коэффициентов аппроксимирующих полиномов

Таким образом, результаты измерений параметров представляются в виде векторов коэффициентов кусочно-полиномиальной аппроксимации и характеризуются ошибкой аппроксимации. Результаты кластеризации векторов коэффициентов с применением алгоритма К-средних показаны на рисунке 6. На рисунке 6а векторы представлены в виде точек в пространстве первой и второй главной компонент, на рисунке 6б приведены статистические характеристики построенных кластеров.



а)

б)

Рисунок 6 – Результаты кластерного анализа параметров ПМГа-ПМГг

Заключение

Описаны метод и модель трансформации, предназначенные для решения задачи обработки данных, поступающих от наблюдаемых технических объектов. При построении процессов учитываются свойства обрабатываемых данных, а также условия их обработки. Построение процессов включает их представление в общем виде и последующую детализацию общего представления до уровня представления исполняемых процессов. Для построения моделей трансформации данных предложены модели обеспечения трансформации. Они представляют собой систему моделей, включающую функциональную, информационную и процессно-сервисную модели. Правила построения моделей определяются в управляющей модели. Модели обеспечения трансформаций представляются в виде онтологических моделей. Онтологические модели строятся на основе классификаторов, описывающих требования, предъявляемые к результатам обработки, обрабатываемые данные, процессы обработки, алгоритмы обработки, а также методы оценки получаемых результатов.

Предложенные модели и методы трансформации позволяют с единых позиций описывать различные процессы обработки данных.

СПИСОК ИСТОЧНИКОВ

- [1] **Llinas, J.** “Revisiting the JDL data fusion model II / Llinas J., Bowman C., Rogova G., Steinberg A., Waltz E., White F. // Technical Report, DTIC Document, 2004.
- [2] **Blasch, E.P.** JDL level 5 fusion model “user refinement” issues and applications in group tracking / Blasch E.P., Plano S. // Proc. of the Signal Processing, Sensor Fusion, and Target Recognition XI, April 2002. P.270–279.
- [3] **Steinberg, A.** Revisions to the JDL Data Fusion Model. Sensor Fusion: Architectures, Algorithms, and Applications / Steinberg A., Bowman C., White F. // Proceedings of the SPIE. 1999. Vol. 3719.
- [4] **Водяхо, А.И.** Архитектурный фреймворк, ориентированный на поддержку процесса разработки многомерных измерений параметров пространственно распределенных объектов. Архитектурный подход к управлению ИТ-сервисами. / А.И. Водяхо, С.С. Голяк, С.А. Гордеев, Н.А. Жукова // Известия СПбГЭТУ «ЛЭТИ» 2013, №4, с. 24-29.
- [5] **Жукова, Н.А.** Принципы организации управления процессами обработки и анализа многомерных измерений в ИГИС / Н.А. Жукова, А.В. Панькин // 5-я Российская мультikonференция по проблемам управления «Информационные технологии в управлении», г. Санкт-Петербург, 9-11 октября, 2012 г. - С. 403-414.
- [6] **Назаров, А.В.** Современная телеметрия в теории и на практике / А.В. Назаров, Г.И. Козырев, И.В. Шитов и др. — СПб.: Наука и техника, 2007. — 672 с.
- [7] **Chang, K.** Principal curves for nonlinear feature extraction and classification / K. Chang, J. Ghosh // Proc. SPIE 3307, Applications of Artificial Neural Networks in Image Processing III, 120 April 1, 1998.
- [8] **Feng, S.** Classification of Curves in 2D and 3D via Affine Integral Signatures / S. Feng, I.A. Kogan, H. Krim // Acta Appl Math (2010) 109: 903. <https://doi.org/10.1007/s10440-008-9353-9>.
- [9] **Hallin, M.** Rank tests for time series analyses: a survey / M. Hallin, M. Puri // In new Directions In Time Series Analyses, Springer-Verlag, New York, 111-154, 1992.
- [10] **Gevins, A.** Statistical pattern recognition / A. Gevins, A. Remond // Handbook of electroencephalography and clinical neurophysiology: methods of brain and magnetic signals, vol 1. Elsevier, Amsterdam. 1987. - Correlation analyses. Handbook of electroencephalography and clinical neurophysiology: methods of brain and magnetic signals, vol 1. Elsevier, Amsterdam. 1987.
- [11] **Бокс, Дж.** Анализ временных рядов прогноз и управление / Дж. Бокс, Г. Дженкинс. Под ред. В.Ф. Писаренко. — М.: Мир, 1974. — Кн. 1. — 406 с. — Кн. 2.— 197 с.
- [12] **Cellucci, C.** Statistical validation of mutual information calculations: Comparison of alternative numerical algorithms / C. Cellucci, A. Albano, P. Rapp // Phys Rev E 71: 066208. 2005.
- [13] **Kugiumtzis, D.** Assessment of Measures of Scalar Time Series Analysis in Discriminating Preictal States / D. Kugiumtzis, I. Vlachos, A. Papana, P.G. Larsson. // International Journal of Bioelectromagnetism, Vol. 9, No 3, p.134-145, 2007.
- [14] **Hinich, M.** Testing for Dependence in the Input to a Linear Time Series Model. Journal of Nonparametric Statistics, 1996.
- [15] **Hinich, M.** Berkeley Detecting Nonlinearity in Time Series: Surrogate and Bootstrap Approaches / M. Hinich, M. Eduardo, E. Mendes, L. Stone // Studies in Nonlinear Dynamics & Econometrics Vol 9, Issue 4 2005 Electronic Press, 2005.
- [16] **Grassberger, P.** Estimation of the Kolmogorov entropy from a chaotic signal / P. Grassberger, I. Procaccia // Phys. Rev. A 28, 2591, 1983.
- [17] **Kugiumtzis, D.** Measures of Analysis of Time Series (MATS): A MATLAB Toolkit for Computation of Multiple Measures on Time Series Data Bases / D. Kugiumtzis, A. Tsimpiris // Journal of Statistical Software. February 2010; 33(5).
- [18] **Дерипаска, А.О.** Адаптивный выбор процессов обработки и анализа многомерных измерений в интеллектуальных информационных системах / А.О. Дерипаска, Н.А. Жукова, А.В. Панькин // Сб. трудов Тринадцатой нац. конф. по искусственному интеллекту с межд. участием. (Белгород, 16–20 окт., 2012 г.). Белгород: Изд-во БГТУ, 2012. Т.1. С. 181–189
- [19] **Городецкий, В.И.** Ассоциативная классификация: аналитический обзор / В.И. Городецкий, О.Н. Тушканова // Труды СПИИРАН. 2015. - Часть 1. Вып. 38. С. 183-203. - Часть 2. Вып. 39. С. 212-240.
- [20] **Tianxing, M.** A Knowledge-based Recommendation System for Time Series Classification / M. Tianxing, N. Zhukova, N. Mustafin // Proceedings of the 24th Conference of Open Innovations Association FRUCT (Moscow, Russia — April 08 - 12, 2019) Article No. 24, 2019. P.751–759.
- [21] The R Project for Statistical Computing. - https://cran.r-project.org/web/packages/available_packages_by_name.html.
- [22] Когнитивные информационные системы мониторинга / А.В. Васильев и др.; Санкт-Петербургский государственный электротехнический университет им. В.И. Ульянова (Ленина) "ЛЭТИ", Национальный

ONTOLOGY MODELS FOR TRANSFORMATION OF DATA ABOUT THE STATE OF TECHNICAL OBJECTS

N.A. Zhukova

*Saint-Petersburg Institute for Information and Automation RAS. Saint Petersburg, Russia
nazhukova@mail.ru*

Abstract

To monitor technical objects many sensors are installed on them, which allows to collect data about these objects. The processes of data processing are determined according to the properties of the collected data and the practical problems that are solved. When data or requirements to the results of its processing change, the processes are rebuilt. To ensure efficient processing, a method and model of data transformation are proposed. They provide a formal description of the sequences of transformations in a general form, as well as their presentation in the form of processing processes. For the practical implementation of the method, it is proposed to represent transformations in the form of a number of linked ontological models, including functional, informational, and process-service models. The construction of models is carried out using rules that are organized in the form of a control model. An ontological model for the transformation of data presented in the form of time series is considered. The structure of its classes is determined by classifiers defined for the initial time series and their representations, as well as methods, algorithms and procedures for their processing and evaluation of results. An example of processing the results of measurements of pressure parameters of a technical object for space purposes during the technical control of its condition is given.

Key words: *transformation models, data processing, time series.*

Citation: *Zhukova AI. Ontology models for transformation of data about the state of technical objects [In Russian]. Ontology of designing. 2019; 9(3): 345-360. – DOI: 10.18287/2223-9537-2019-9-3-345-360.*

References

- [1] *Llinas J, Bowman C, Rogova G, Steinberg A, Waltz E, White F.* Revisiting the JDL data fusion model II, Technical Report, DTIC Document, 2004.
- [2] *Blasch EP, Plano S.* JDL level 5 fusion model “user refinement” issues and applications in group tracking. Proceedings of the Signal Processing, Sensor Fusion, and Target Recognition XI, April 2002. P.270–279.
- [3] *Steinberg A, Bowman C, White F.* Revisions to the JDL Data Fusion Model. Sensor Fusion: Architectures, Algorithms, and Applications // Proceedings of the SPIE. 1999. Vol. 3719.
- [4] *Vodyakho AI, Golyak SS, Gordeyev SA, Zhukova NA.* An architectural framework focused on supporting the process of developing multidimensional measurements of parameters of spatially distributed objects. An architectural approach to managing IT services. Izvestiya SPbGETU «LETI». 2013; 4: 24-29.
- [5] *Zhukova NA, Pan'kin AV.* Principles of the organization of process control and analysis of multidimensional measurements in IGIS // 5th Russian Multiconference on Management Problems “Information Technologies in Management” (ITU-2012) , St. Petersburg, October 9 - 11, 2012. - P. 403-414.
- [6] *Nazarov AV, Kozurev GI, Shitov IV and others.* Modern telemetry in theory and practice. — SPb.: Nayka i Tekhnika, 2007. — P. 672.
- [7] *Chang K, Ghosh J.* Principal curves for nonlinear feature extraction and classification Proc. SPIE 3307, Applications of Artificial Neural Networks in Image Processing III, 120 April 1, 1998
- [8] *Feng S, Kogan IA, Krim H.* Classification of Curves in 2D and 3D via Affine Integral Signatures. Acta Appl Math (2010) 109: 903. <https://doi.org/10.1007/s10440-008-9353-9>.
- [9] *Hallin M, Puri M.* Rank tests for time series analyses: a survey. In new Directions In Time Series Analyses , Springer-Verlag, New York, 111-154. 1992.
- [10] *Gevins A, Remond A.* Statistical pattern recognition. Handbook of electroencephalography and clinical neurophysiology: methods of brain and magnetic signals, vol 1. Elsevier, Amstrdam. 1987. - Correlation analyses. Handbook of

electroencephalography and clinical neurophysiology: methods of brain and magnetic signals, vol 1. Elsevier, Amsterdam. 1987.

- [11] **Box G, Jenkins G.** Time Series Analysis Forecast and Management. Holden-Day, Inc. San Francisco, CA, USA 1990. ISBN:0816211043.
 - [12] **Cellucci C, Albano A, Rapp P.** Statistical validation of mutual information calculations: Comparison of alternative numerical algorithms. Phys Rev E 71: 066208. 2005
 - [13] **Kugiumtzis D, Vlachos I, Papana A, Larsson PG.** Assessment of Measures of Scalar Time Series Analysis in Discriminating Preictal States. International Journal of Bioelectromagnetism, Vol. 9, No 3, pp. 134-145, 2007.
 - [14] **Hinich M.** Testing for Dependence in the Input to a Linear Time Series Model. Journal of Nonparametric Statistics, 1996
 - [15] **Hinich M, Eduardo M, Mendes E, Stone L.** Detecting Nonlinearity in Time Series: Surrogate and Bootstrap Approaches. Studies in Nonlinear Dynamics & Econometrics Vol 9, Issue 4 2005 Electronic Press, 2005
 - [16] **Grassberger P, Procaccia I.** Estimation of the Kolmogorov entropy from a chaotic signal. Phys. Rev. A 28, 2591, 1983
 - [17] **Kugiumtzis D, Tsimpiris A.** Measures of Analysis of Time Series (MATS): A MATLAB Toolkit for Computation of Multiple Measures on Time Series Data Bases. Journal of Statistical Software. February 2010, Vol33, Issue 5.
 - [18] **Deripaska AO, Zhukova NA, Pankin AV.** Adaptive selection of processes for multi dimensional measurements processing and analyses in intelligent information systems [In Russian].. Proceedings of 13th national conference on artificial intelligence. (Pelgorod, 16–20 oct., 2012 г.).Belgorod: BGTU, 2012. T.1. P. 181–189
 - [19] **Gorodetskiy VI, Tushkanova ON.** Associative classification: analytical review. Proceedings of SPIIRAS. - Part 1 2015; 38: 183-203. - Part 2. 2015; 39: 212-240.
 - [20] **Tianxing M, Zhukova N, Mustafin N.** A Knowledge-based Recommendation System for Time Series Classification // Proceedings of the 24th Conference of Open Innovations Association FRUCT (Moscow, Russia — April 08 - 12, 2019) 2019; 24: 751–759.
 - [21] The R Project for Statistical Computing. - https://cran.r-project.org/web/packages/available_packages_by_name.html.
 - [22] Cognitive information systems for monitoring / A.V. Vasil'yev and others; Saint Petersburg Electrotechnical University "LETI", Almazov National Medical Research Centre. - Spb.: ETU "LETI", 2017. – 200 p.
-

Сведения об авторах



Жукова Наталья Александровна, 1983 г. рождения. Окончила Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина) (СПбГЭТУ ЛЭТИ) в 2006 году, к.т.н. (2008). Старший научный сотрудник лаборатории вычислительной техники и технологий программирования Санкт-Петербургского института информатики и автоматизации РАН, доцент кафедры математического обеспечения ЭВМ СПбГЭТУ ЛЭТИ и кафедры информатики и прикладной математики Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики. В списке научных трудов более 20 работ в области обработки и анализа данных.

Natalya Alexandrovna Zhukova (b. 1983) graduated from the Saint Petersburg Electrotechnical University "LETI" (ETU) in 2006. Ph.D (2008), researcher at Saint-Petersburg Institute for Information and Automation RAS, associate professor at ETU and ITMO University. She is co-author of more than 20 publications in the field of data processing and analyses.