

6. Kaufman L., Rousseeuw P. J. Finding groups in data: An introduction to cluster analysis. New York, NY, USA: John Wiley & Sons, 2005.

7. Kim J.-O., Mueller C. W. Factor analysis: Statistical methods and practical issues. USA: Sage Publications, Newbury Park, Calif, 1978.

Yu. T. Lyachek, M. C. A. Muthanna, S. S. S. Nasser
Saint Petersburg Electrotechnical University

M. M. A. Muthanna
St. Petersburg University «SPBU»

DENSITY AND DISTRIBUTION OF LORA BASIC STATIONS IN THE «SMART CITY»

This work considers two main challenges associated with one of the most common and recent used IoT network, which is Low Power Wide Area Networks (LPWAN). The first challenge is the energy constraints associated with the design of and the problem of massive traffic results from the huge number of deployed devices. This work presents two main clustering algorithms developed for LoRa based LPWAN. One clustering algorithm considers the base stations and gateways distribution through the network in a way that achieves the best coverage and highest system reliability. This clustering algorithm is static. The second clustering algorithm considers clustering of distributed devices related to each base station to achieve higher energy and latency efficiency. This algorithm is dynamic. Both algorithms are mainly considered for dense LoRa based IoT networks (i. g. smart city networks). The proposed algorithms are based on the two common clustering methods; FOREL and K-means. The proposed algorithms are implemented, for dense IoT network with various intensities, for performance evaluation.

Internet of things, dense networks, LPWAN, LoRa, gateway, clustering, energy

УДК 004.514

Е. С. Новикова, Я. А. Бекенева, М. П. Бестужев
Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В. И. Ульянова (Ленина)

Методы визуального анализа для мониторинга данных от сенсорных сетей

Исследуются подходы к визуальному анализу потоков разнородных данных в режиме реального времени, позволяющие формировать паттерны нормального поведения объекта и выявлять возможные отклонения в нем. В результате анализа определены основные модели визуализации, применяемые для мониторинга потоков данных и выявления аномалий в них, показаны их основные достоинства и ограничения; сформулированы основные требования к проектированию моделей визуализации, применяемых для графического представления разнородных данных в режиме реального времени. Предложен подход к построению аналитической панели мониторинга потоков данных на основе карт плотности, матричного представления и линейных диаграмм, показано ее применение на примере мониторинга данных от системы тепловентиляции и кондиционирования «умного» здания.

Мониторинг в реальном времени, визуализация потока данных, карты плотности, RadViz, матричная визуализация, линейные графики с временной шкалой

Современные системы мониторинга часто состоят из большого количества разнородных устройств, генерирующих потоки данных о разных параметрах контролируемых процессов. Подобная организация систем мониторинга позволяет получить достаточно полное представление о ходе процессов, но при этом возрастает

сложность обработки и анализа разнородных данных. К основным задачам систем мониторинга относится обнаружение внештатных ситуаций, угрожающих действий или другого аномального поведения. Выявление аномалий является сложной аналитической задачей, которая обычно не может быть решена с помощью чисто

визуального анализа или чисто автоматических вычислительных методов. Еще одна сложная задача – это представление оператору результатов обработки и анализа больших объемов данных, в частности, когда работа осуществляется в режиме реального времени и решение должно быть принято незамедлительно. Визуализация данных облегчает анализ и оценку аномалий при использовании различных методов взаимодействия и интуитивно понятного представления контекста, что дает дополнительные доказательства для подтверждения или опровержения выдвинутых аналитиком гипотез [1], [2].

Однако успех применения методов визуальной аналитики для формирования паттернов и обнаружения отклонений в потоке данных зависит от выбранных визуальных моделей и методов автоматического обнаружения аномалий. Согласно [3], большинство методов анализа и визуализации фокусируются на исследовании статических наборов данных, и добавлении динамического компонента в источник данных приводит к серьезным проблемам как для автоматического, так и для визуального анализа данных. Кроме того, современные системы мониторинга подразумевают работу с большими данными [4], что предъявляет особые требования к средствам не только обработки и анализа, но и представления результатов аналитику. Операторы систем мониторинга реального времени должны получать визуальную информацию в удобном для восприятия виде, который позволял бы незамедлительно отреагировать на малейшие отклонения, но в то же время не создавал излишнюю когнитивную нагрузку. Тем не менее, при необходимости аналитик должен получить более подробную информацию для принятия последующего решения.

В статье авторы обсуждают существующие подходы к методам выявления аномалий в потоках данных в режиме реального времени, базирующиеся на методиках визуализации; основное внимание уделяется используемым моделям графического представления данных, обсуждаются их преимущества и недостатки, на их основе формулируются требования к моделям визуализации многомерных данных, применяемых для мониторинга объектов в режиме реального времени. Представлен прототип панели управления для мониторинга многомерных потоков данных на основе карт плотности, матричного представления и линейных диаграмм, протестированный

для мониторинга системы тепловентиляции и кондиционирования здания.

Данная статья является расширенной версией тезисов, опубликованных на конференции ElConRus2020 [5]. Она отличается расширенным анализом работ, оценкой достоинств и недостатков существующих подходов к визуальному анализу потоков данных, выявленными требованиями к проектированию методик визуализации и более подробным описанием предложенного подхода.

Подходы к анализу потоков данных сенсорных сетей в режиме реального времени. Анализ потока данных от сети сенсоров методами визуальной аналитики можно рассматривать как задачу визуализации временных рядов, однако необходимо помнить, что требование обеспечения анализа в реальном времени накладывает определенные ограничения на используемые модели визуализации, поскольку оператор должен постоянно уделять внимание контролируемым переменным и быть готовым отреагировать на изменения в любой момент времени.

Традиционно для мониторинга потоков данных применяется множество взаимосвязанных линейных графиков с осью времени [6]. Например, для мониторинга сетевой активности в системах защиты информации и управления событиями используются линейные диаграммы [7], [8]. К. Маткович и др. адаптировали интерфейс стрелочных приборов для мониторинга потоков данных [9]. Это позволило авторам компактно показать текущие значения параметров в контексте с историческими данными, таким образом, любое отклонение в потоке данных становится заметно аналитику. Однако при использовании такого способа представления данных необходимо тщательно продумывать сценарии их использования для обеспечения четкости и читаемости значений контролируемых параметров. Кроме того, широко распространены мнемосхемы объектов – планы этажей зданий, функциональные схемы объектов и приборов – поскольку именно мнемосхема отображает реальные связи между параметрами и компонентами объекта, подлежащего управлению или контролю. Они обычно используются для отображения текущих значений параметров, контроль за их изменением осуществляется с помощью линейных графиков с временной шкалой.

М. К. Хао и др. исследовали возможности матричного представления информации для визу-

ализации потоковых данных [10]. Каждая ячейка таблицы представляет собой усредненное значение параметра за некоторой фиксированный период времени, например за одну минуту. Существуют различные варианты заполнения таблицы, например, новые значения могут добавляться по столбцам, т. е. сначала заполняется текущий столбец, а затем новый, или по строкам, когда сначала заполняются строки. В [10] авторы исследовали три различные схемы обновления таблицы, которые отличаются способом смещения данных и их наложением. Они показали, что для каждого метода необходимо найти компромисс между последовательностью отображения данных, частотой их обновления и потенциальным непониманием оператора. В [11] предложенная модель визуализации была использована для обнаружения аномалий: если значение элемента данных превышает некоторый порог, заданный пользователем, то соответствующая ячейка таблицы отмечается небольшим прямоугольником. Аналогичный подход был реализован в [12] для мониторинга данных по добыче нефти. В [13] представлена дискретно-вероятностная модель матричной визуализации, позволяющей строить как двумерные, так и 3D-изображения.

В общем случае поток событий распределен во времени неравномерно, некоторые периоды времени перегружены различными событиями, в то время как другие почти «пусты». При использовании обычной временной шкалы возникает проблема наложения множества событий, которая может скрыть данные, важные для анализа. Отображение усредненных характеристик событий может скрыть важные изменения в данных. Для решения этой проблемы Ф. Фишер и соавторы ввели понятие *расслабленной временной оси* (relaxed time axis) [14]. Временная шкала представлена последовательностью временных интервалов с заданной длительностью. Каждое событие кодируется цветным прямоугольником, цвет которого зависит от атрибутов события, например от уровня его критичности. Когда в течение некоторого интервала времени фиксируется большое число различных событий, ширина каждого пря-

моугольника, соответствующего событию, уменьшается, чтобы «освободить место» для вновь добавленных событий. В результате использования расслабленной временной оси информация об абсолютном времени наступления события теряется, однако аналитик по-прежнему располагает информацией о порядке их наступления, поскольку более поздние события располагаются правее предыдущих. Авторы показали, что такое решение позволяет сравнивать различные потоки данных и выявлять определенные тренды в них. Схожий принцип организации временной шкалы, отражающий порядок событий, используется в [15] для представления событий, которые характеризуют поведение пользователя в информационной системе. Действие пользователя в системе представлено цветным прямоугольником, цвет которого зависит от типа этого действия. На рис. 1 показаны различные сеансы пользователя, представленные последовательностью действий с учетом их порядка во времени. Часто повторяющиеся последовательности выделены увеличенным размером соответствующих графических элементов, что позволяет аналитикам быстро обнаруживать типичное поведение пользователя, а также сравнивать между различными сессиями.

В некоторых случаях для отражения циклического характера данных используются модели визуализации со шкалой времени, расположенной по спирали или по окружности по аналогии с часами [16], [17]. Графические элементы, обозначающие данные, располагаются по спирали, номинальные данные часто отображаются с помощью простых иконок, а количественные данные – с помощью цвета, толщины линии или текстуры. Например, в [17] для мониторинга сетевого трафика в режиме реального времени предложена модель визуализации *Traffic Circle*, ось времени в которой располагается по спирали, а сетевые потоки отображаются в виде дуг, длина которых зависит от продолжительности сетевого сеанса, а положения начала и конца дуги на окружности определяются временем (порядком) появления/завершения потока. Г. Джанетско и др. адаптировали похожую модель визуализации для монито-

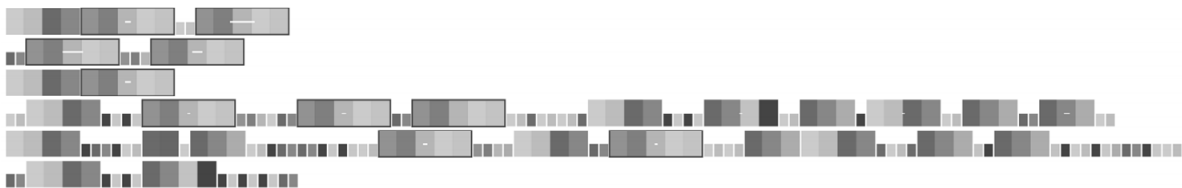


Рис. 1

ринга расхода электропотребления и выявления аномалий в энергетической системе [18]. Значения контролируемого параметра кодируются с помощью цвета пикселей, которые располагаются по окружности.

Ф. Фишер и Д. Кайм предложили концепцию срезов данных для мониторинга данных из разных источников [19]. Его основная идея заключается в объединении элементов данных в срезы данных в режиме реального времени, если изменения в них не превышают некоторый заданный порог. Срезы данных отображаются в виде множества моделей графического представления, объединенных визуально. Данные из разных источников представлены различными методами визуализации – например, текстовые данные представлены с использованием облака слов, структуры графов – с использованием древовидных карт или связанных графов. Цвет фона срезов данных отражает уровень сходства между ними. Предложенная концепция была реализована авторами в инструменте NStreamAware, апробированном для мониторинга от датчиков кибербезопасности и событий от социальных сетей.

Интересный подход к мониторингу и анализу данных представлен в [20]. В его основе лежит попарная оценка подобия характера изменения данных сенсоров, полученные оценки позволяют выполнить проекцию сенсоров в зависимости от того, как меняются их показатели, на двухмерную плоскость и отобразить их. Таким образом, каждый сенсор представляет собой точку на плоскости, причем точки находятся рядом, если изменение их данных похоже. Если характер данных от сенсора сильно изменяется, то меняется положение соот-

ветствующего сенсора на плоскости, а его траектория выделяется цветом (рис. 2). Очевидно, что данный подход может быть применен для оценки и мониторинга данных от сенсоров одного типа.

В основе подхода, представленного в [21], лежат марковские логические сети. Основным назначением среды служит выявление нарушений в процессах, при этом возможна одновременная визуализация нескольких процессов, однако источники данных для всех исследуемых процессов должны быть однородными.

В [19] методика проецирования потока многомерных событий в двумерное пространство адаптирована для мониторинга сетевого трафика хостов компьютерной сети. В качестве методики проецирования исходного пространства в двумерное используется алгоритм RadViz [20], который также позволяет сгруппировать объекты с одинаковым поведением и выявить потенциальные аномальные отклонения в них.

Таким образом, анализ подходов, представленных в литературе, показал, что для визуализации потоков данных используются следующие методики: 1) линейные графики с временной шкалой [4]–[6]; 2) мнемосхемы; 3) пиксель-ориентированные методики/матричное представление данных [7]–[9], [12]–[14]; 4) методики на основе диаграмм Ганта, отражающих последовательность событий/действий [10], [11]. Выбор этих методик объясняется, в первую очередь, их простотой – они могут быть использованы даже лицами, не имеющими операторского опыта. Мнемосхемы дают информацию непосредственно о структуре объекта, функциональных связях, что формирует предметный контекст анализа, однако отслеживание

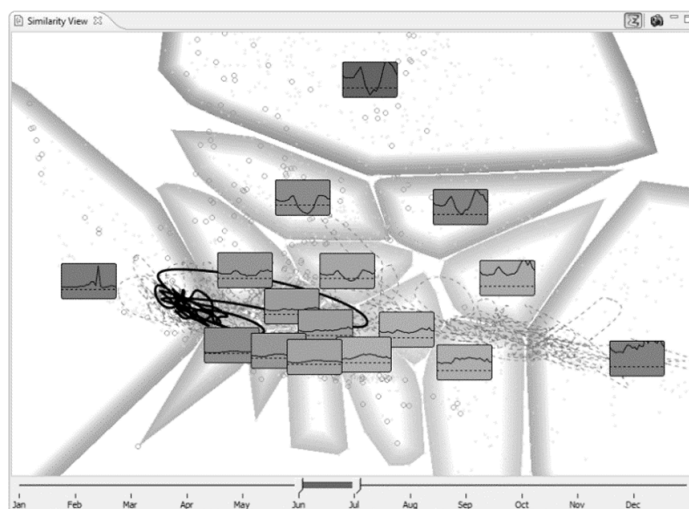


Рис. 2

трендов в изменении данных и формирование паттернов поведения объектов с их помощью затруднены. Линейные диаграммы чаще всего используются для отображения значений определенного параметра, они позволяют легко увидеть как аномалии, так и тренды в данных, однако совместный анализ нескольких параметров уже достаточно сложен. Пиксель-ориентированные методики визуализации и матричное представление данных дают возможность отобразить большие объемы данных компактно, что, с одной стороны, позволяет выявлять паттерны функционирования объекта/объектов и аномалии в них, а с другой стороны, поддерживает процесс визуальной корреляции данных, т. е. определение возможных взаимосвязей между значениями контролируемых параметров с помощью их графического представления [21]. Последнее является важной особенностью данных методик визуализации, поскольку в настоящее время потоки данных поступают от множества различных источников, и для формирования ситуационной осведомленности в режиме реального времени необходимо иметь возможность оценивать множество параметров одновременно.

Отдельно стоит отметить модели визуализации, в основе которых лежат методики проецирования многомерного пространства на двумерную плоскость [22]. Они не столь распространены на практике, однако исследования показали, что их применение также позволяет формировать графические паттерны нормального поведения объектов, обладающие высокой способностью снижения размерности входного потока данных.

Подход к визуальному анализу потока данных от разнородных источников. На основе выполненного анализа представленных в открытой литературе методик визуального анализа потока данных от разнородных источников были сформированы следующие требования к их проектированию:

- 1) обеспечение контекста анализа и ментальной карты объекта мониторинга;
- 2) привлечение внимания аналитика к оповещениям (аномалиям);
- 3) возможность отображения данных от разных источников;
- 4) уменьшение когнитивной нагрузки на аналитика.

Авторы предлагают следующий подход к мониторингу потока событий от различных устройств, который отвечает вышеперечисленным требованиям. При проектировании данного подхода авторы придерживались принципа поиска информации, сформулированного Б. Шнейдерманом: «Сначала обзор – фильтр и масштабирование – детали по запросу» [23]. Очевидно, что для анализа данных в режиме реального времени этот принцип можно переформулировать следующим образом: «Контекст анализа – показ важного – детали по требованию». В данном случае под контекстом анализа авторы понимают паттерны нормального и аномального поведения, построенные на основе собранных ранее данных или сформированные на основе описания особенностей функционирования системы, если такая информация доступна проектировщику. «Показ важного» связан с необходимостью снизить когнитивную нагрузку на оператора и уменьшить объем отображаемых данных. Возможность получить детальную информацию по изменениям параметров связана с возникающей необходимостью объяснить наблюдаемые отклонения в функционировании системы на этапе «показ важного». В режиме реального времени аналитику требуется достаточно быстро увидеть изменения в ходе процесса, которые могли бы указывать на аномалию, однако при этом требуется показать значения наиболее важных характеристик [24].

В основе предлагаемого подхода лежит предположение, что состояние системы в некоторый момент времени может быть представлено точкой многомерного пространства, положение которой определяется множеством значений, контролируемых параметрами, служащими координатными осями данного многомерного пространства. Идея представить таким образом систему для проектирования моделей визуализации была предложена в [25], где авторы предложили ряд методик для исследования набора данных от системы тепло-вентиляции и кондиционирования. Несмотря на то что источники данных разнородны, формат данных схож, а исходные данные, используемые в данной публикации, приведены к единому формату. В связи с этим можно утверждать, что предлагаемый подход одинаково применим как к однородным данным, так и к данным от разнородных источников, но имеющим схожий формат или же прошедшим некоторую предварительную обработку.

В предлагаемом подходе ключевыми элементами являются: 1) карта плотности, построенная на основе результатов проецирования многомерного пространства в двумерное («контекст анализа»); 2) матричное представление контролируемых параметров («показ важного»); 3) линейные графики («детали по требованию»).

Карта плотности построена на основе визуализации RadViz – многомерного алгоритма визуализации данных, который выполняет проекцию исходного пространства в 2-мерное. В [26] показано, что применение алгоритма RadViz позволяет строить графические паттерны нормального и аномального функционирования системы, причем для нормальных режимов функционирования характерны плотные скопления точек, аномалии же отличаются сильным разбросом точек. Эти результаты позволили авторам использовать оценку плотности распределения точек на плоскости как признак нормального функционирования системы, а карту плотности, построенную на основе результатов RadViz-визуализации данных, – как механизм отображения типичных режимов функционирования системы. Таким образом, если точка, соответствующая текущему состоянию системы, лежит в области с высокой плотностью, то это состояние системы нормально, но если она находится в области с низкой плотностью, это может быть признаком возможной аномалии. Траектория точки может быть использована для прогнозирования ее состояния. В настоящем подходе для построения карты плотности используется ядерное сглаживание оценки плотности распределения точек на плоскости [27].

Матричное представление данных используется для отображения изменений в потоке данных. Авторы считают, что необходимо привлечь внимание оператора к таким изменениям, однако не все их необходимо контролировать, поскольку причиной многих изменений в показаниях датчиков служат незначительные шумы. Другой важный момент, который требуется учесть при проектировании модели визуализации, – это то, что длительность некоторых подозрительных изменений может быть чрезвычайно мала. По этим причинам авторы приняли решение использовать концепцию срезов данных для их агрегирования в режиме реального времени, если изменения в них не превышают некоторого заданного оператором порога. Таким образом, оператор определяет до-

пустимые отклонения, в пределах которых состояние системы считается неизменным. Такое решение позволяет значительно снизить объем отображаемых данных, а также уменьшить когнитивную нагрузку на оператора, так как отрисовка изменений осуществляется, только если превышен порог допустимого отклонения, а не на постоянной основе.

Для формирования срезов данных применяется модифицированный алгоритм пространственной кластеризации dbscan [28]. Его оригинальная версия предназначена для анализа статистических данных, поэтому авторам потребовалось внести изменения в его работу для учета времени генерации точек, описывающих состояние системы в некоторый момент времени, и переопределить базовые точки алгоритма через граничные точки множества. Параметр ϵ алгоритма dbscan задает значение порога, определяющее допустимое отклонение точки, принадлежащей срезу данных, от базовых точек. Если оно для новой точки не превышает заданного порога, то она добавляется в текущий срез данных, в противном случае создается новый срез данных, в который входит эта точка. Визуализация срезов данных осуществляется с помощью матрицы, строки которой соответствуют отслеживаемым параметрам, а столбцы представляют собой срезы данных. Каждая ячейка отображает среднее значение атрибута в срезе данных. Насыщенность цвета используется для кодирования значений. Чем выше значение, тем темнее ячейка матрицы.

Очевидно, что матричное представление данных дает информацию только о характере их изменения, поэтому для отображения реальных значений параметров используются традиционные линейные диаграммы с временной шкалой. Пользователь имеет возможность выбрать исследуемые параметры потока данных, чтобы отслеживать значения в режиме реального времени.

Предложенный подход был протестирован на данных, моделирующих функционирование системы отопления, вентиляции и кондиционирования [29]. Функционирование этой системы описывается множеством датчиков, постоянно отслеживающих температурный режим здания и выдающих управляющие сигналы на специальное кондиционирующее и отопительное оборудование, поэтому ее функционирование легко может быть представлено в виде траектории точки в

многомерном пространстве. На рис. 3 представлен программный прототип разработанной панели мониторинга параметров системы тепловентиляции, общих для всего здания. Карта плотности с нормальным режимом функционирования системы находится в верхнем левом углу панели управления (рис. 3А). Она обновляется по мере того как формируется новый срез данных, т. е. только тогда, когда изменение какого-либо параметра превышает установленное его значение. Текущее состояние системы отображается точкой на карте плотности и взаимосвязано с матричным представлением (рис. 3В). Линейные графики, представляющие собой привычный элемент для операторов, занимают центральное положение на панели управления (рис. 3С). Детали в табличном виде отображаются при нажатии на элемент тепловой карты, линейного графика и матричного представления.

Выполненное первичное тестирование пользовательского опыта показало, что предложенные модели визуализации достаточно понятны. Карта плотности наглядно отображает поведение системы, однако не дает представления о деталях. Матричное представление было отнесено к методам визуализации, которые достаточно быстры для понимания, но все же предоставляемый ими уровень детализации недостаточен. Однако данный недостаток легко компенсируется применением традиционных линейных графиков. Следует отметить, что именно матричное представление позволяет эффективно обнаруживать аномальные отклонения и значительно сокращает объемы отслеживаемых данных. Некоторые трудности воз-

никли из-за нелинейного преобразования отображаемой временной шкалы в матричном представлении, поскольку размер матричного элемента не зависит от длительности среза данных. Это означает, что необходимо изменить алгоритм компоновки столбцов матрицы, чтобы выделить текущий элемент. Возможным решением может стать увеличение последнего столбца, чтобы заполнить пространство дисплея.

Авторы планируют также усовершенствовать прототип программного обеспечения посредством разработки методов взаимодействия, более тесно связывающих визуализацию на основе матрицы и линейные диаграммы, оценить быстродействие системы мониторинга и минимальные системные требования к вычислительным ресурсам. Планируется разработать рекомендации по организации рабочего процесса анализа и выбора порога для создания срезов данных. Другое направление работ связано с тестированием подхода для анализа потоков разнородных данных, имеющих принципиально разные форматы. Поскольку современные системы мониторинга зачастую состоят из множества таких устройств, как различные датчики, измерительные системы, а также системы фото- или видеорегистрации, следует разрабатывать методы, которые позволяли бы осуществлять совместный визуальный анализ данных от таких источников. В качестве дальнейших этапов исследования и разработки планируется реализовать анализ данных от других систем, например от систем водочистных сооружений, поскольку выбранные методики визуализации достаточно универсальны и не зависят от исследуемой предметной области.

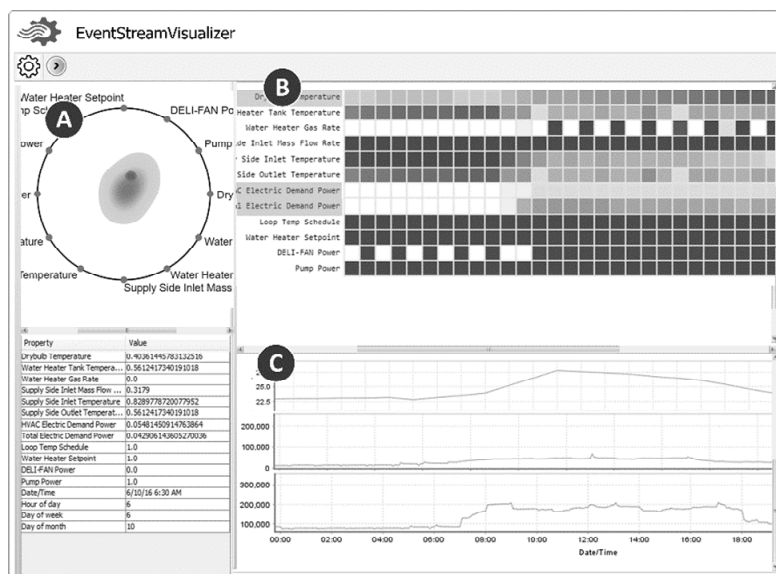


Рис. 3

В статье были рассмотрены подходы к визуальному анализу многомерных потоков данных. Авторы показали, что к наиболее часто используемым методикам визуализации относятся линейные графики с временной шкалой, мнемосхемы, пиксель-ориентированные и матричные способы представления, а также методики на основе диаграмм Ганта. Особый интерес также представляют методики, основанные на проецировании многомерного пространства в двумерное. В работе был предложен подход к мониторингу потока данных от разных источников, в основе которого лежит идея представления элемента потока данных в виде точки в многомерном пространстве. Эта идея определила выбранные визуальные мо-

дели и алгоритмы для агрегирования данных, однако она также подразумевает определенные ограничения на практическое использование панели мониторинга, поскольку атрибуты потока данных должны быть числового типа. Подход был протестирован на наборе данных, описывающих функционирование системы отопления, вентиляции и кондиционирования умного здания. Эксперименты показали, что панель управления позволяет определить типичное функционирование системы и выявить аномальные отклонения. Дальнейшие направления работы посвящены совершенствованию прототипа и оценке удобства использования предлагаемой визуально-аналитической системы.

СПИСОК ЛИТЕРАТУРЫ

1. Keim D. A., Mansmann F., Thomas J. Visual analytics, How much visualization and how much analytics // ACM SIGKDD Explorations. 2009. № 11(2). P. 5–8.
2. Visual analytics for event detection: Focusing on fraud / R. A. Leite, T. Gschwandtner, S. Miksch, E. Gstrein, J. Kuntner // Visual Informatics. 2018. № 2(4). P. 198–212.
3. Mansmann F., Fischer F., Keim D. A. Dynamic visual analytics – Facing the real-time challenge. Expanding the Frontiers of Visual Analytics and Visualization. London: Springer, 2012. P. 69–80.
4. Zeng Y. R., Chang Y. S., Fang Y. H. Data Visualization for Air Quality Analysis on Bigdata Platform // Intern. Conf. on System Science and Engineering (ICSSE). Russia, Novosibirsk: IEEE, 2019. P. 313–317.
5. Bestuzhev M. P., Novikova E. S., Bekeneva Y. A. Analysis of the visual analytics techniques for monitoring heterogeneous data streams // Proc. of IEEE Conf. of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus). St. Petersburg and Moscow, 2020. P. 237–240.
6. Visualization of Time-Oriented Data / W. Aigner, S. Miksch, H. Schumann, C. Tominski // London: Springer Science & Business Media, 2011.
7. Novikova E., Kotenko I. Analytical visualization techniques for security information and event management // Proc. of the 21st Euromicro Intern. Conf. on Parallel, Distributed and Network-based Processing (PDP 2013), Belfast, Northern Ireland. Los Alamitos, California, 2013. P. 519–525.
8. Singh K., Best P. Anti-money laundering: Using data visualization to identify suspicious activity // Intern. J. of Accounting Information Systems. 2019. № 34. P. 100418.
9. Process visualization with levels of detail / K. Matkovic, H. Hauser, R. Sainitzer, M. E. Groller // Proc. of the IEEE Symp. on Information Visualization. Boston, 2002. P. 67–70.
10. Density displays for data stream monitoring / M. C. Hao, D. A. Keim, U. Dayal, D. Oelke, C. Tremblay // Computer Graphics Forum. 2008. № 27(3). P. 895–902.
11. Visual analytics of anomaly detection in large data streams / M. C. Hao, U. Dayal, D. A. Keim, R. K. Sharma, A. Mehta // Proc. of the Visualization and Data Analysis Conf. San Jose, 2009. № 7243. P. 72430B.
12. An application of sensor and streaming analytics to oil production / K. Viswanathan, C. Gupta, C. Lakshminarayan, M. C. Hao, U. Dayal, R. Vennelakanti, P. Helm, S. Rangaiah, H. R. Sagiraju, S. Doddmani // Proc. of the 17th Intern. Conf. on Management of Data. Bangalore, India, 2011. № 18. P. 1–8.
13. Preuss D., Weinkauff T., Kruger J. A discrete probabilistic approach to dense flow visualization // IEEE Transactions on Visualization and Computer Graphics, 2020 Jul. 13. URL: <https://arxiv.org/abs/2007.01629> (дата обращения 23.07.2020).
14. Fischer F., Mansmann F., Keim D. A. Real-time visual analytics for event data streams // In Proc. of the 27th Annual ACM Symp. on Applied Computing. New York, 2012. P. 801–806.
15. A visual analytics approach for user behaviour understanding through action sequence analysis / P. Nguyen, C. Turkay, G. Andrienko, N. Andrienko, O. Thonnard // Proc. of EuroVis Workshop on Visual Analytics. Barcelona, 2017. P. 67–71.
16. Weber M., Alexa M., Müller W. Visualizing time-series on spirals // Proc. of the IEEE Symp. on Information Visualization (InfoVis). San Diego, Ca, USA, 2001. P. 7–13.
17. Real-time visualization of network behaviors for situational awareness / D. M. Best, S. Bohn, D. Love, A. Wynne, W. A. Pike // Proc. of the Seventh Intern. Symp. on Visualization for Cyber Security (VizSec '10). New York, NY, USA: ACM, 2010. P. 79–90.

18. Anomaly detection for visual analytics of power consumption data / H. Janetzko, F. Stoffel, S. Mittelstädt, D. A. Keim // *Computer & Graphics*. 2014. Vol. 38. P. 27–37.

19. Fischer F., Stoffel F. NStreamAware: Real-Time visual analytics for data streams (VAST Challenge 2014 MC3) // *IEEE Conf. on Visual Analytics Science and Technology*. Paris, 2014. P. 373–374.

20. Visual analysis of time-series similarities for anomaly detection in sensor networks / M. Steiger, J. Bernard, S. Mittelstädt, H. Lücke-Tieke, D. Keim, T. May, J. Kohlhammer // *Computer Graphics Forum*. 2014. № 33. P. 401–410.

21. Formalizing visualization design knowledge as constraints: Actionable and extensible models in draco / D. Moritz, C. Wang, G. L. Nelson, H. Lin, A. M. Smith, B. Howe, J. Heer // *IEEE transactions on visualization and computer graphics*. 2018. № 25(1). P. 438–448.

22. Mansmann F., Meier L., Keim D. A. Visualization of host behavior for network security // *Proc. of the Workshop on Visualization for Computer Security (VizSEC 2007)*. Sacramento, California, USA, 2007. P. 187–202.

23. Ankerst, M., Berchtold, S., Keim, D. A. Similarity clustering of dimensions for an enhanced visualization of multidimensional data // *IEEE Symp. on Information*

Visualization (INFOVIS '98). Washington: IEEE Computer Society, 1998. P. 52–60.

24. Shneiderman B. Dynamic queries for visual information seeking // *IEEE Software*. 1994. Vol. 11, № 6. P. 70–77.

25. Visual analytics for domain experts: Challenges and lessons learned / C. Arbesser, T. Mühlbacher, S. Kormornyik, H. Piringer. 2017. URL: <https://www.vrvis.at/publications/pdfs/PB-VRVis-2017-019.pdf> (дата обращения 10.10.2020).

26. Novikova E., Bestuzhev M., Kotenko I. Anomaly detection in the HVAC system operation by a RadViz based visualization-driven approach. *Computer security // Lecture Notes in Computer Science*. 2019. Vol. 11980. P. 402–418.

27. Härdle W. *Applied nonparametric regression*. Cambridge: University Press, 1990.

28. A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise / M. Ester, H. P. Kriegel, J. Sander, X. Xu // *Proc. of the Second Intern. Conf. on Knowledge Discovery and Data Mining*. Portland, 1996. P. 226–231.

29. Vast Challenge Homepage. URL: <http://vacomunity.org> (дата обращения 5.03.2020).

E. S. Novikova, Ya. A. Bekeneva, M. P. Bestuzhev
Saint Petersburg Electrotechnical University

VISUAL ANALYTICS METHODS FOR MONITORING HETEROGENEOUS DATA FLOWS

Studies existing approaches to visual analysis of heterogeneous data streams in real time, allowing to form patterns of normal behavior of an object and identify possible deviations in it. The paper outlines the main visualization models used to monitor data flows and identify anomalies in them are identified, their main advantages and limitations are shown; the basic requirements for the design of visualization models used to graphically represent heterogeneous data in real time are formulated. An approach to the construction of an analytical panel for monitoring data flows based on density maps, matrix representations, and linear diagrams is proposed; its application is shown on the example of monitoring data from a smart building ventilation and air conditioning system.

Real-time monitoring, data stream visualization, density maps, RadViz, matrix visualization, time line charts
