УДК 004.056

И. Н. Муренин, Е. С. Новикова
Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В. И. Ульянова (Ленина)

Визуализация аномальной активности в перемещениях сотрудников критической инфраструктуры

Представлен подход к анализу перемещений сотрудников критической инфраструктуры, отличающийся сочетанием алгоритмов интеллектуального анализа данных интерактивных методик визуализации. Поиск групп пользователей со схожим поведением осуществляется с помощью самоорганизующихся сетей Кохонена, результат функционирования которых настраивается с помощью разработанной авторами модели визуализации BandView. Для обнаружения аномалий в поведении сотрудников предлагается механизм оценки поведенческих отклонений, использующий значения их пространственно-временных атрибутов. Подход апробирован на наборе данных, предоставленном в рамках конкурса VASTMiniChallenge-2 2016, который описывает перемещения сотрудников по зданию организации.

Выявление аномалий, визуальная аналитика, тепловые карты, самоорганизующиеся карты, паттерны поведения, оценка отклонений

В настоящее время данные, описывающие перемещения движущихся точек в пространстве, являются наиболее распространенным типом пространственных данных, и их анализ имеет множество практических применений. Такие наборы данных обычно называют траекториями. Анализ траекторий позволяет создавать шаблоны поведения объектов для последующего обнаружения возможных аномалий в изучаемых наборах данных. В этом случае результаты исследования траекторий могут быть использованы для обеспечения безопасности движения воздушного, наземного и водного транспорта посредством мониторинга местоположения, траектории, скорости транспортного средства и обнаружения неожиданных препятствий на маршруте [1]–[4]. Мониторинг перемещений сотрудников критических инфраструктур обеспечивает контроль соблюдения мер и политик безопасности, установленных на предприятии, а также способствует выявлению внутренних нарушителей безопасности [5].

В статье описываются подход к обнаружению пространственно-временных аномалий в перемещениях персонала организации, отличающийся сочетанием автоматических методик анализа данных и графического представления данных. Множество интерактивных моделей визуализации не только осуществляет отображение выявленных шаблонов в перемещениях сотрудников и потенциально аномальных ситуаций, но и позволяет контролировать результат применения автоматических методов анализа данных. Паттерны поведения представляются с помощью разработанной авторами модели визуализации, отражающей последовательность контролируемых зон, посещаемых сотрудником, и длительность пребывания в них. Отклонения в перемещениях сотрудников отображаются в виде тепловой карты. Для уменьшения возможных шумов на тепловой карте разработан механизм оценки отклонений, позволяющий сфокусироваться на отдельных подозрительных выбросах. Он оценивает аномалии на основании пространственно-временных атрибутов отклонений, таких, как место, продолжительность и время. Данная методика анализа маршрутов сотрудников *отличается* от представленных в литературе подходов [6]–[8] анализом пространственно-временных характеристик перемещений в контексте атрибутов контролируемых зон; использованием интерактивных самоорганизующихся карт Кохонена для выявления групп сотрудников, обладающих схожим поведением; оценкой аномалий с учетом периодичности появления отклонений в маршруте сотрудника.

Подход протестирован на тестовых данных, предложенных в рамках конкурса VastChallenge 2016[1].

[1] Vast Challenge Website [Online], Available:http://vacommunity.org/.

Подход к обнаружению аномалий в перемещениях пользователей. В основе предложенного подхода лежат сеть SOM, используемая для выявления групп сотрудников с похожими шаблонами перемещений, и тепловые карты, позволяющие обнаруживать периоды аномальной активности. Эти модели визуализации дополняются вложенными графиками, которые отражают последовательность посещения контролируемых зон сотрудником, и таблицей свойств объектов, выбранных аналитиком. Все элементы графического представления данных интерактивны и логически взаимосвязаны между собой. Например, при выборе элемента сети SOM обновляются данные таблицы свойств, тепловой карты и вложенных графиков.

Исходные данные. В представленном подходе исходными данными являются журналы, формируемые считывателями контролируемых зон и содержащие информацию только о времени посещения контролируемой зоны определенным сотрудником. Кроме того, доступна информация о позициях большинства сотрудников внутри организации, а также карта ее здания. Отличительной особенностью логов, формируемых считывателями, является то, что регистрация сотрудников происходит непосредственно в момент их появления в контролируемой зоне. Таким образом, появление записей в журнале логов носит нерегулярный характер, и интервал между ними для определенного сотрудника может длиться от нескольких секунд до нескольких часов. Также стоит заметить, что некоторые сотрудники могут совершать множество перемещений по зданию организации в соответствии с их должностными обязанностями, в то время как другие значительно реже покидают свои рабочие места. Таким образом, длины временных рядов, описывающих перемещения отдельных сотрудников, могут заметно варьироваться. В предложенном подходе исходные временные ряды преобразуются в векторы фиксированной длины.

Введем следующие обозначения. Пусть $E = \{e_i\}_{i=1}^{n}$ – множество сотрудников, $Z = \{z_j\}_{j=1}^{m}$ – множество контролируемых зон, $T = \{t_k : t_i < t_j; i < j\}_{k=1}^{p}$ – упорядоченное множество временных меток. Соответственно, записи из журнала считывателей контролируемых зон имеют вид (e_i, z_j, t_k). Общий интервал времени, представленный первой и последней записями журнала, обозначим как T_0, а множество записей обозначим как $\text{LOGS} = \{(e_i, z_j, t_k)\}$, $i = 1, ..., n$, $j = 1, ..., m$, $k = 1, ..., p$. Интервал времени T_0 разбивается на последовательность одинаковых временных интервалов Δt, т. е.

$$T_0 = \{\Delta t_l : \Delta t_i = \Delta t_j; \Delta t_i = [t_i; t_{i+1});$$

$$\Delta t_{i+1} = [t_{i+1}; t_{i+2}); i \neq j; i, j \leq l\}_{l=1}^{r}.$$

На каждом интервале времени Δt_l для каждого сотрудника определяется количество посещений $n_{z_j}^{\Delta t_l}$ и продолжительность пребывания $\Delta t_{z_j}^{\Delta t_l}$ в каждой контролируемой зоне z_j. Таким образом, множество LOGS можно представить в виде множества упорядоченных пар $\text{LOGS} = \{(n_{z_j}^{\Delta t_l}; \Delta t_{z_j}^{\Delta t_l})\}_{e_i}$, $i = 1, ..., n$, $j = 1, ..., m$, $l = 1, ..., r$, вычисленных для каждого сотрудника e_i $(i = 1, ..., n)$. Именно эта последовательность формирует множество атрибутов траектории пользователя, совмещая пространственно-временные характеристики движения.

В предлагаемом подходе длительность интервала Δt_l составляет 4 ч. Эксперименты показали, что этого достаточно для обнаружения даже незначительных временных отклонений длиной в 5 мин.

Модель визуализации на основе сети SOM. Для обнаружения групп сотрудников с похожим поведением и индивидуальных шаблонов поведения для отдельных сотрудников используется нейронная сеть SOM (и ее графическое представление – SOM-карта). Ее основная задача заключается в выполнении разведочного анализа данных при отсутствии информации об их структуре. Сеть SOM является искусственной нейронной сетью, обучаемой без учителя. Она состоит из узлов, или нейронов, которым соответствуют векторы весовых коэффициентов с размерностью, определяемой размерностью исходных данных. В ходе итерационного обучения входные векторы сравниваются с весовыми векторами каждого нейрона, а веса нейрона, наиболее соответствующие входному вектору, и веса соседних нейронов подстраиваются таким образом, чтобы быть ближе к входному вектору. Для корректного обучения сети SOM исходные данные должны быть целостны и не иметь пропусков отдельных значений для каждого атрибута. Однако в рассматриваемом случае этап предва-

рительной обработки данных гарантирует получение векторов с явно определенными значениями для каждого атрибута.

Нейронные сети SOM являются также методикой визуализации многомерных данных, осуществляющей графическое отображение данных в двумерное пространство. Для представления структуры сети SOM часто используется U-матрица [9]. Она отражает структуру данных с помощью отображения среднего расстояния между весовыми векторами соседних нейронов. Чем темнее цвет узла, тем дальше он находится от своих соседей. Узлы, окрашенными цветами с похожей яркостью, схожи друг с другом. Узлы, содержащие центры кластеров, отмечены кругом, размер которого пропорционален количеству объектов в кластере.

В описываемом подходе нейронная сеть SOM используется для обнаружения групп сотрудников с похожим поведением. Вектор признаков сотрудника описывает его активность на протяжении всего исследуемого периода времени. Это позволяет предположить, что однократные или редкие отклонения в перемещениях не влияют на результат кластеризации. Благодаря этому карта SOM отображает различия в траекториях сотрудников в соответствии с особенностями их ролей в организации.

Точность кластеризации траекторий сотрудников зависит от числа узлов сети SOM. Чем больше размер сети, тем больше формируемых кластеров. В этом случае узлы сети SOM, соответствующие работникам со схожим поведением, будут сосредоточены в регионе SOM, раскрашенном схожими оттенками. Эксперименты показали, что разница в поведении сотрудников, принадлежащих к соседним узлам, незначительна и объясняется, как правило, небольшими отклонениями продолжительности пребывания в пределах контролируемой зоны, в то время как число посещенных зон и даже последовательности посещений сохраняются. Такие различия во времени редко превышают 5…10 мин, поэтому есть возможность использовать одну общую пространственно-временную траекторию для всех сотрудников, принадлежащих к одному региону SOM. Для объединения таких сотрудников в один кластер рекомендуется настраивать размер карты SOM в соответствии с их количеством.

Модель визуализации BandView. Модель визуализации BandView предназначена для выявления связи между пространственными и временными атрибутами перемещений. Она представляет собой вложенную столбиковую диаграмму, горизонтальная ось которой обозначает время. По вертикальной оси откладываются или сотрудники организации, или дни, если отображаются маршруты одного сотрудника. Маршрут работника отображается в виде последовательности блоков-сегментов, каждый из которых соответствует временному интервалу, в течение которого работник находится в данной зоне. Цвет сегмента используется для кодирования расположения зоны. Цветовая схема кодирования контролируемых зон построена таким образом, что каждому этажу соответствует определенный цвет, а цветовая палитра для зон, расположенных на этом этаже, создается изменением яркости соответствующего цвета: чем больше значение идентификатора контролируемой зоны – тем темнее будет цвет соответствующего сегмента. На рис. 1 показаны траектории, пройденные сотрудниками одной должности в течение одного дня. Зоны первого этажа отображаются в темно-серых оттенках, зоны второго этажа – в светло-серых, а зоны третьего – в серых. Из рисунка следует, что большинство сотрудников проводят рабочий день на втором этаже, и только один человек работает в офисе на третьем.

Модель BandView используется при отображении шаблонов и аномалий, а также исходных данных. Модель визуализации BandView позволяет обнаружить, где и когда произошла аномалия, как долго она длилась, поскольку обеспечивает визуальное сравнение маршрута сотрудника с маршрутами его коллег. Для работы с исходными данными предложен гибкий механизм фильтрации, позволяющий строить сложные логические выражения, используя всевозможные атрибуты перемещений: идентификатор сотрудника, должность, офис, продолжительность пребывания в зоне, идентификатор зоны, этаж и т. д.

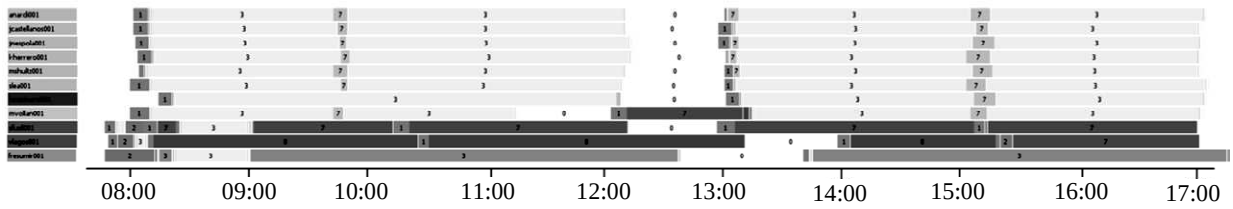

Рис. 1

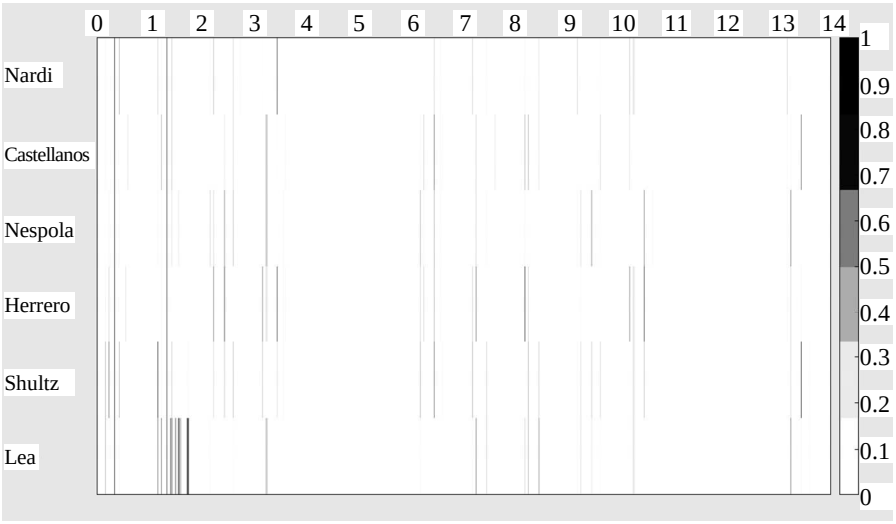

Рис. 2

Отображение аномалий с помощью тепловых карт и механизм оценки аномалий. Цель тепловой карты – отразить наличие потенциально аномальных отклонений в движении персонала. Авторы считают, что аномалии проявляются в нерегулярных нечастых изменениях в поведении субъекта. Таким образом, предлагается искать отклонения в рамках группы сотрудников или группы дней, имеющих схожие шаблоны перемещений. Тепловая карта строится следующим образом: по оси Y откладываются сотрудники, принадлежащие одному кластеру, а по оси X – атрибуты векторов признаков для данных сотрудников. Каждый элемент тепловой карты представляет собой расстояние от значения атрибута вектора признаков сотрудника до центроида соответствующего ему кластера. Для того чтобы закодировать дистанцию от кластера или оценку отклонения, используется цвет. Темные цвета на тепловой карте соответствуют большим значениям параметра.

На рис. 2 изображена тепловая карта отклонений в перемещениях группы сотрудников, принадлежащих одному отделу и обладающих одинаковым поведением.

Отображение расстояний без предварительной обработки может привести к возникновению зашумленной картинки в случае, если расстояния от центроида до объектов внутри кластера примерно одинаковы, а также к исчезновению отклонений на карте, если разница между отдельными расстояниями слишком велика. Для решения этих проблем разработан механизм оценки отклонений, рассматривающий их в контексте некоторого среднего значения для времени пребывания сотрудника в определенной зоне в течение определенного промежутка времени.

Механизм оценки состоит из двух составных частей – расчета значимости (рейтинга) отклонения и определения пороговых значений для каждой контролируемой зоны.

Оценка значимости отклонения осуществляется с помощью z-показателя (z-score), который отражает расстояние данного значения от среднего значения по набору данных. Для использования этого показателя делается предположение, что исходное распределение отклонений в передвижениях сотрудников на заданном множестве интервалов времени Δt_i можно считать нормальным с параметрами $\left(\overline{(e_i; z_j)}, \sqrt{nD(z_j, \Delta t_i)}\right)$, где n – количество дней, описываемых логами электронных считывателей. Дискретизация данных на этапе подготовки позволяет сформировать выборку интервалов времени Δt_i, которая характеризуют активность сотрудника на интервале времени с учетом некоторого заданного периода времени (сутки, неделя и т. д.). Благодаря этому имеется возможность оценить отклонения в поведении сотрудников на интервале с 8 ч утра до 12 ч для каждого рабочего дня.

Пороговые значения для каждой зоны непосредственно определяют, какие отклонения будут отображаться на тепловой карте, а какие – нет. Пороги определяются аналитиком для каждой зоны в зависимости от ее характеристик. Например, можно предположить, что отклонение в 10 % от правой границы доверительного интервала для средней продолжительности пребывания в кафе является допустимым в обеденное время и аномальным во второй половине рабочего дня. Таким образом пороги позволяют удалить не представляющие интереса шумы с тепловой карты.

Эксперимент и обсуждение результатов. Для оценки предложенного подхода был использован набор данных, представленный в рамках конкурса TheVASTChallenge 2016: Mini-Challenge 2. Он содержит журнал логов считывателей контролируемых зон, регистрирующих появление пользователя в контролируемых зонах здания организации. Когда сотрудник при помощи прокси-карты попадает в новую подконтрольную зону, данные о его карте автоматически распознаются датчиком и записываются в журнал. Следует отметить, что большая часть зон доступна для сотрудников, даже если они забыли свои пропуска. Журнал содержит записи датчиков за две недели. Аналитик имеет в своем распоряжении схему планировки здания и расположения офисов, включая карты подконтрольных зон, а также список сотрудников с указанием их должностей и офисов.

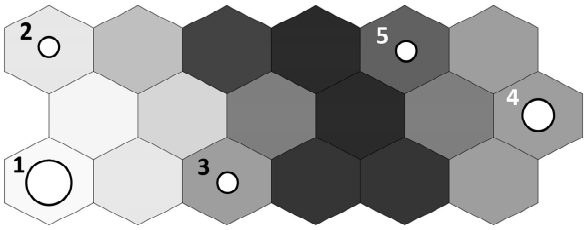

Рис. 3

На первом этапе анализа было обнаружено, что большинство сотрудников одного отдела передвигаются одинаково, и их маршруты могут незначительно отличаться из-за расположения рабочего места или некоторых особенностей функциональных обязанностей. На рис. 3 дана SOM-карта, на которой представлены 5 кластеров сотрудников, принадлежащих отделу безопасности. Типичные маршруты их передвижений показаны на рис. 1. Члены наиболее многочисленной группы № 1 имеют лишь 2 периодических шаблона поведения: один для рабочих дней, другой – для выходных. Их рабочий день обычно начинается около 8:00 и заканчивается в 17:00. Около 12:00 они выходят на обед из здания на час. Члены этой группы проводят большую часть своего рабочего времени в зоне 2–3, расположенной на втором этаже, в которой согласно плану находятся их офисы. Каждые полтора-два часа они покидают свою зону и переходят в прилегающую зону 2–7 на 5…10 мин. В зоне 2–7 находятся офисы и туалетные комнаты. В выходные дни члены группы не приходят на работу. Два ближайших к этой группе кластера № 2 и 3 состоят из одного человека. Их траектории в какой-то степени схожи с передвижением сотрудников из первой группы, так как их офисы расположены в зоне 2–3, где они также проводят большую часть рабочего времени. Например, поведение сотрудника, принадлежащего кластеру № 2, незначительно отличается от поведения сотрудников группы № 1. Основное отличие заключается в том, что он совершает обход зоны 2–7 только во второй половине дня около 15.00. Исследование его перемещений с помощью модели визуализации BandView и тепловой карты позволило выявить небольшое отклонение, произошедшее на третий день всего контролируемого периода. В этот день во время послеобеденного обхода зоны 2–7 он не возвращается в свою комнату, а в 17:00 идет домой из зоны 2–7, что можно увидеть на рис. 4. Это отклонение можно объяснить тем, что сотрудник мог забыть где-то свой пропуск. Оставшиеся две группы сотрудников № 4, 5 отделены от других в карте SOM поло-

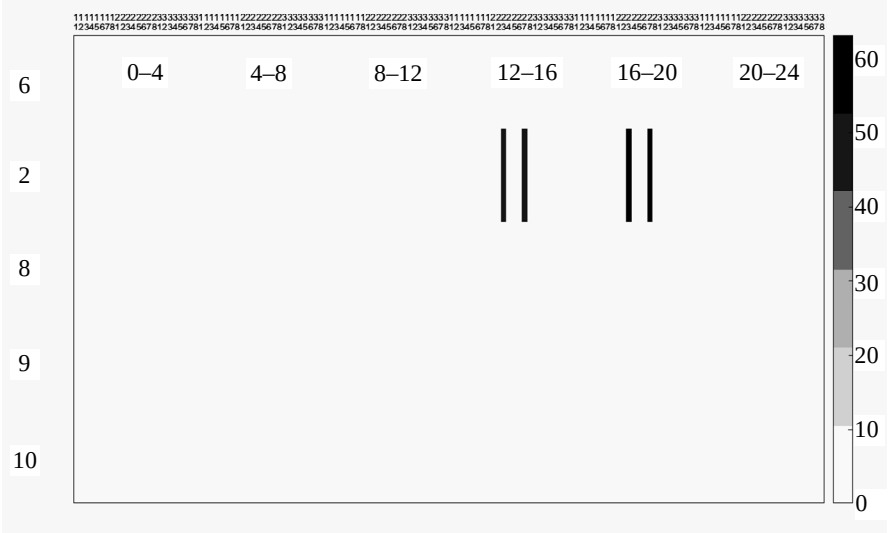

Рис. 4

сой темных ячеек. Это позволяет предположить, что их маршруты значительно отличаются от траекторий, описанных ранее. Кластер № 4 состоит из двух сотрудников, чьи рабочие места располагаются на первом этаже. Они имеют 2 периодичные модели перемещений, зависящие от дня недели. Один из них посещает зону 2–3 каждый вторник в первую половину рабочего дня, в то время как другой посещает эту же зону в первую половину дня каждого четверга. Кластер № 5 состоит из одного сотрудника. Благодаря модели BandView становится понятно, что он обычно находится на третьем этаже. Тот факт, что на третьем этаже располагаются офисы руководства компании, позволяет авторам предположить, что этот сотрудник является начальником отдела безопасности. Его перемещения достаточно разнообразны и сильно зависят от дня недели.

Эксперименты показали, что анализ тепловой карты в сочетании с моделью визуализации BandView позволяет достаточно легко найти отклонения следующего вида: посещение зоны в необычное время; отсутствие сотрудника на рабочем месте в положенное время; дублирование логов электронных считывателей; случаи, когда сотрудник возвращается на свое рабочее место или забывает использовать пропуск при выходе из здания. Следует отметить, что длительность обнаруженных отклонений может варьироваться от десятков секунд до нескольких часов.

В большинстве случаев анализ взаимодействия персонала улучшает понимание причин возникновения аномалий. Модель BandView может быть полезна для понимания возможного взаимодействия между работниками. Однако она работает лишь в том случае, когда число отображаемых работников ограничено и не превышает 10–15 человек, иначе отслеживание взаимоотношений между сотрудниками сильно затруднено. Таким образом, одно из основных направлений будущей работы связано с разработкой динамической методики визуального анализа, позволяющей исследовать взаимодействия между сотрудниками. Другое направление будущей работы затрагивает анализ данных, полученных из различных источников. Журналы операционной системы, например события входа-выхода и события клавиатуры, являются доказательствами того, что сотрудник находится на рабочем месте. Показания сенсоров инфраструктур здания, например, таких, как тепловентиляционная система, могут также объяснить нетипичное поведение сотрудников. Корреляция этих данных требует тщательной разработки новых методов визуального анализа с учетом особенностей исходных журналов.

В статье предлагается подход для формирования пространственно-временных шаблонов и обнаружения аномалий в перемещениях сотрудников критических инфраструктур, основанный на использовании методик визуального анализа. Ключевые элементы подхода – интерактивная SOM-карта, использующаяся для обнаружения групп сотрудников, и тепловая карта, модифицированная рейтинговыми и пороговыми механизмами для оценки отклонений в контексте шаблона поведения сотрудника. Для иллюстрации предложенного подхода был использован набор данных, предоставленный в рамках конкурса VastChallenge 2016. В статье обсуждены полученные результаты и определены направления будущей работы, посвященной совершенствованию прототипа, развитию методик визуализации и оценки эффективности предлагаемой аналитической системы визуализации.

Работа выполнена при финансовой поддержке РФФИ № 16-07-00625.

СПИСОК ЛИТЕРАТУРЫ

1. Millonig A., Maierbrugger G. Identifying unusual pedestrian movement behavior in public transport infrastructures // Proc. of Movement Pattern Analysis Workshop (MPA2010), Zurich, 2010. P. 106–110.

2. The use of Bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the Ghent Festivities / M. Versichele, T. Neutens, M. Delafontaine, N. Van de Weghe // App. Geography. 2012. Vol. 32(2). P. 208–220.

3. Lerman Y., Rofe Y., Omer I. Using Space Syntax to Model Pedestrian Movement in Urban Transportation Planning // Geographical Analysis. 2014. Vol. 46(4). P. 392–410.

4. Analysis and visualisation of movement: an interdisciplinary review / U. Demšar, K. Buchin, F. Cagnacci, K. Safi, B. Speckmann, N. Van de Weghe, R. Weibel // Movement Ecology. 2015. Vol. 3(1). P. 5.

5. A multi-agent based framework for the simulation of human and social behaviors during emergency evacuations / X. Pan, C. Han, K. Dauber, K. Law // AI & Society. 2007. Vol. 22. P.113–132.

6. Legg P. A. Visualizing the Insider Threat: Challenges and tools for identifying malicious user activity // Proc. of the 2015 IEEE Symp. On Visualization For Cyber Security (VizSec-2015), Chicago, 2015. P. 1–7.

7. Guo C. Dodeca-Rings Map: Interactively Finding Patterns and Events in Large Geo-temporal Data // IEEE Symp. on Visual Analytics Science and Technology (VAST). 2014. P. 353–355.

8. Schreck J. Bernard T. Von Landesberger J. Kohlhammer. Visual cluster analysis of trajectory data with interactive Kohonen maps // Information Visualization. 2009. Vol. 8, № 1. P. 14–29.

9. Ultsch A. Self-organizing neural networks for visualization and classification // Information and Classification. P. 307–313. DOI: 10.1007/978-3-642-50974-2_31.

I. N. Murenin, E. S. Novikova
Saint Petersburg Electrotechnical University «LETI»

VIS UALIZATIONOF THE ANOMALOUS ACTIVITYIN TRAJECTORIES OF THE EMPLOYEES' OF THE CRITICAL INFRASTRUCTURE

Thepaperpresentsanapproachtoanalysisofthemovementsofthecriticalinfrastructurestuff characterized by usage of the data mining algorithms and interactive visualization techniques. The groups of the employees with similar behavior are determined using Kohonen self-organizing maps, which are set up using special BandView visualization model designed by the authors. To detect anomalies in employees' behavior, a special mechanism for rating deviations based on assessment of their spatiotemporal attributes is proposed. The approach his test edusing data set given within VAST MiniChallenge-2 2016, which describes employees' movement in the organization building.

Anomaly detection, visual analytics, heat maps, self-organizing maps, behavior patterns, anomaly rating

УДК 681.32

К. А. Кноп, С. Ю. Лузин, М. С. Лузин, С. А. Сорокин
ООО «ЭРЕМЕКС» (Санкт-Петербург)

Ю. Т. Лячек
Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В. И. Ульянова (Ленина)

Расстановка фанаутов в области BGA с нерегулярным расположением контактов

Рассмотрены проблемы расстановки переходных межслойных отверстий (фанаутов) на печатной плате в области BGA-компонентов. Предложен алгоритм поиска места фанаута для каждой контактной площадки в области BGA, основанный на размещении в местах, свободных от контактов, максимального числа переходов, решении задачи о назначениях и удалении избыточных переходов. Алгоритм реализован в САПР TopoR и позволяет получить результаты лучшие, чем при использовании аналогичных процедур в других САПР.

Печатный монтаж, автоматическая трассировка, BGA-компоненты, размещение фанаутов

Автоматическая трассировка проводников на печатной плате (ПП) при традиционном ортогональном проектировании [1] осуществляется последовательно, при этом не контролируется наличие ресурсов площади платы для прокладки еще не проложенных проводников. Подобная стратегия может приводить к блокировке контактных площадок на слое.

На рис. 1 показана ситуация, когда прокладка от крайних контактов проводников в противоположных направлениях блокирует возможность подхода на слое к остальным контактам компонента.

Рис. 1

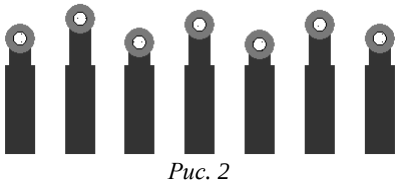

Рис. 2