Применение учетных баз данных при формировании информации о событиях на производственных объектах

Я. А. Бекенева
Санкт-Петербургский государственный электротехнический университет
«ЛЭТИ» им. В.И. Ульянова (Ленина)
Санкт-Петербург, Россия
yabekeneva@etu.ru

***Аннотация*. Мониторинг производственных объектов осуществляется с помощью различных специализированных устройств, в том числе с помощью различных систем контроля доступа. Существует необходимость проведения анализа данных, поступающих от систем мониторинга, с целью выявления различного рода нарушений. При формировании информации о событии, а также выявлении нарушений, существует ряд задач, связанных с выявлением несоответствий между объектами наблюдения или режимных нарушений. Выявление таких нарушений невозможно без использования дополнительных учетных систем, в которых хранится информация о графиках работы сотрудников, закрепленного за ними оборудования и т.д. В работе представлен подход, позволяющий определить порядок использования учетных систем при проведении анализа данных от мониторинговых систем для выявления нарушений регламента.**

Ключевые слова: анализ данных; системы мониторинга; источники данных; учетные базы данных

I. Введение

Анализ данных от различных источников мониторинга с целью выявления возможных аномалий или серьезных нарушений является актуальной задачей, так как неуклонно повышаются требования к безопасности и противодействию терроризму, предупреждения порчи и краж имущества организаций и т.д. Современные системы мониторинга содержат различные устройства, осуществляющие контроль действий, исполняемых в рамках производственных процессов. Ряд исследований посвящен выявлению аномалий в перемещениях транспортных средств [1] на территории производственной организации, перемещениях сотрудников на территории офисного здания [2], движении грузовых составов [3]. Существующие способы выявления аномалий основаны на применении таких методов как классификация, кластеризация, построение прогностических моделей.

При подготовке данных актуальной является задача корреляции данных от разнородных устройств контроля. Одно событие может быть одновременно зарегистрировано несколькими различными устройствами, но описывать событие с помощью разного состава атрибутов. При этом существует проблема неопределенностей, связанная с недостатком данных, позволяющих однозначно отнести отдельно взятую запись к какому-либо событию. В простейшем случае к одному событию относятся все записи, для которых совпадают временные и пространственные атрибуты, а также атрибуты объекта наблюдения [4]. Однако в ряде случаев не все устройства контроля предоставляют информацию об объекте наблюдения. В случае анализа перемещений транспортных средств разные системы контроля доступа и фото- или видеонаблюдения могут предоставлять информацию как о транспортном средстве, так и о водителе, что затрудняет задачу объединения данных на основе идентификатора объекта наблюдения.

Некоторые нарушения могут также быть связаны с нарушением режима (например, присутствие сотрудника на рабочем месте в непредусмотренное расписанием время).

Для решения подобных задач могу быть использованы дополнительные источники данных, которыми могут являться учетные базы данных. Такие базы данных могут содержать информацию о графике работы сотрудников, графике отпусков, соответствии сотруднику закрепленного за ним транспортного средства или оборудования и т.п. Информация из таких баз, как правило, предназначена для хранения, учета и планирования рабочего процесса, однако не используется для анализа. При этом использование такой информации может позволить как облегчить задачу объединения разнородных данных о событии, так и выявить нарушения, которые могли бы быть пропущены в ином случае.

II. Описание подхода

A. *Установление соответствия объектов наблюдения с помощью учетных баз данных*

Основная идея подхода может быть проиллюстрирована на примере установления соответствия транспортных средств, используемых для осуществления перевозок на территории предприятия, и

Работа выполнена при финансовой поддержке Стипендии Президента Российской Федерации СП-2581.2019.5

водителей, управляющих ими. Разнородные устройства контроля позволяют идентифицировать и зарегистрировать разные объекты наблюдения. Например, с помощью средств визуального мониторинга могут быть идентифицированы номера транспортных средств, а с помощью систем контроля доступа – номера пропусков, выписанные на фамилии водителей. В таком случае может быть однозначно объединена информация о событии, в которой в качестве зарегистрированного объекта наблюдения выступает водитель или транспортное средство. Однако существует задача объединения таких записей для наиболее полного описания события. В простейшем случае в одном месте в одно время происходит одно событие, и две группы записей могут быть объединены в одно событие лишь на основе пространственно-временных факторов.

Однако предлагаемый подход позволяет решить одновременно две задачи:

- заполнение пропущенных атрибутов на основе учетной информации;
- выявление возможных нарушений уже на этапе подготовки данных.

Эти задачи могут быть проиллюстрированы на примерах. На рис. 1 представлен фрагмент данных, полученных от разнородных источников и прошедших этап первичного объединения в единую таблицу. В качестве атрибутов идентификатора объекта наблюдения выступают номер транспортного средства (VNumber) и фамилия водителя (Name). Данные об одном событии регистрируются одновременно тремя устройствами контроля, где два устройства регистрируют номер транспортного средства, а одно устройство регистрирует фамилию водителя. Таким образом, одному событию соответствуют три строки в данной таблице. Для проведения полноценного анализа данных необходимо объединение данных о событии в одну строку, содержащую всю полезную информацию от всех имеющихся устройств контроля. При этом возможно однозначное объединение лишь двух строк, для которых заполнен номер транспортного средства.

VNumber	Name	Date	Ty
Р678КА		11.06.2018	
М674ЫЛ		11.06.2018	
Д348МП		11.06.2018	
Д348МП		11.06.2018	
	Назаров	11.06.2018	
	Назаров	11.06.2018	
М674ЫЛ		11.06.2018	
	Ильин	11.06.2018	
	Ильин	11.06.2018	
	Булычев	11.06.2018	
Р678КА		11.06.2018	
М674ЫЛ		11.06.2018	
М674ЫЛ		11.06.2018	
Р678КА		11.06.2018	
Р678КА		11.06.2018	
Р678КА		12.06.2018	
	Ильин	12.06.2018	
	Ильин	12.06.2018	
М674ЫЛ		12.06.2018	

Рис. 1. Фрагмент исходного файла

На рис. 2 представлена база данных, содержащая информацию о соответствии фамилий водителей и транспортных средств. В простейшем случае такая база может содержать записи с однозначным соответствием фамилии человека и закрепленного за ним транспортного средства. В приведенном примере использование транспортных средств осуществляется разными водителями в зависимости от графика их работы и поставленных задач по транспортировке грузов.

VNumber	Name	Beg date	End date
Р678КА	Назаров	11.06.2018	12.06.2018
М674ЫЛ	Ильин	11.06.2018	12.06.2018
Д348МП	Булычев	11.06.2018	11.06.2018
Т483АС	Ильин	16.06.2018	18.06.2018
М674ЫЛ	Михайло	17.06.2018	18.06.2018
Р678КА	Булычев	13.06.2018	13.06.2018
М674ЫЛ	Щедрин	13.06.2018	13.06.2018
Д348МП	Булычев	15.06.2018	15.06.2018
А158КМ	Щедрин	14.06.2018	15.06.2018
А158КМ	Булычев	16.06.2018	18.06.2018

Рис. 2. Пример учетной базы данных

Специально разработанный плагин на языке Java позволил реализовать основную идею подхода и осуществить взаимное заполнение атрибутов для последующего объединения данных о событии. Программа позволяет извлечь из основного файла идентификатор объекта наблюдения и дату, когда было зафиксировано событие, осуществить вложенный поиск в базе данных и найти соответствующий идентификатор объекта другого типа. Таким образом, после выполнения плагина полученный файл не содержит пропущенных значений атрибутов, идентифицирующих объект наблюдения. Дальнейшая интеграция данных об одном событии может быть осуществлена с использованием любого поля, как фамилии водителя, так и номера транспортного средства.

При подобном взаимном автозаполнении идентификаторов объектов наблюдения могут быть выявлены нарушения, при которых водитель управляет не тем транспортным средством, которое закреплено за ним на данный период времени. Такие аномалии могут быть обнаружены в связи с невозможностью объединить три записи о событии в одну, так как идентификаторы объектов будут заполнены в соответствии с базой данных, и заполненные автоматически значения атрибутов будут не совпадать.

Так как в приведенном примере все события совершаются в зонах, оснащенных тремя устройствами контроля, то при объединении записей об одном событии количество строк в таблице уменьшится ровно в три раза. Если в полученном после преобразования файле окажется большее количество строк, это может свидетельствовать об аномальных ситуациях, при этом неполные события могут быть легко обнаружены в полученной таблице.

B. Выявление нарушений регламента без использования взаимного автозаполнения атрибутов

Предлагаемый подход учитывает возможность выявления нарушений регламента без использования плагина для взаимного заполнения атрибутов. В таком случае объединение данных о событии осуществляется на основании пространственно-временных атрибутов события, и по умолчанию считается, что при отсутствии двух одновременных событий в одном месте проверка атрибутов объекта наблюдения является излишней.

Выходной файл представляет собой уменьшенный в три раза по сравнению с исходным файлом, где три записи об одном событии объединены в одну. При этом идентификаторы объектов наблюдения заполнены в соответствии с их значением в соответствующих строках.

Далее для каждой строки проверяется дата, идентификаторы объектов и сравниваются с информацией, хранящейся в базе. При соответствии объектов наблюдения учетной информации событие считается легитимным по данному признаку. При выявлении несоответствия событие считается подозрительным, и тщательной проверке подлежит вся последовательность событий для каждого из указанных объектов наблюдения.

III. Заключение

Использование учетных баз данных в процессах преобразования и анализа данных от систем мониторинга может позволить повысить точность проводимого анализа и выявить аномалии, которые было бы сложно или невозможно выявить без использования таких баз. В то же время базы, содержащие графики дежурств при сменном режиме, графики отпусков позволили бы не только выявить нелегитимные посещения или легитимные непосещения сотрудниками предприятия. Учетные базы данных могут позволить оценить события, связанные с нахождением на рабочем месте сотрудников, находящихся в отпуске или на выходных, а также непосещения ими рабочего места в рабочие дни. Такие и подобные им нарушения могут быть выявлены как при использовании средств визуализации для анализа данных [5], так и различных методов интеллектуального анализа данных [6].

В дальнейших работах планируется более глубокое изучение вопросов, затронутых в данной работе. Планируется также исследование использования учетных баз для преобразования данных для последующего применения к ним методов интеллектуального анализа процессов.

Список литературы

[1] Абрамов Н.А. Компьютеризированная система контроля трафика на крупных предприятиях (система «Цербер») / Н.А. Абрамов // Труды Института системного анализа Российской академии наук. 2012. Т. 62. №. 3. С. 3-10.

[2] Novikova E., Bekeneva Y., Shorov A. The Motif-Based Approach to the Analysis of the Employee Trajectories within Organization //Security and Communication Networks. 2018. Т. 2018.

[3] Шабельников А.Н. Интеллектуальный метод предсказания появления нештатных ситуаций в процессе расформирования поездов на сортировочной горке / А.Н. Шабельников, А.В. Суханов, С.М. Ковалев // Инженерный вестник Дона. 2015. Т. 38. №. 4. С.1-26.

[4] Бекенева Я.А. Преобразование данных от разнородных систем мониторинга //Программные продукты и системы. 2019. Т. 32. №. 2.

[5] Kotenko I., Novikova E. Vissecanalyzer: A visual analytics tool for network security assessment //International Conference on Availability, Reliability, and Security. Springer, Berlin, Heidelberg, 2013. С. 345-360.

[6] Kholod I.I., Efimova M., Rukavitsyn A., Shorov A. Time Series Distributed Analysis in IoT with ETL and Data Mining Technologies // Internet of Things, Smart Spaces, and Next Generation Networks and Systems. Springer, Cham, 2017. С. 97-108.