

АНАЛИЗ И ОБРАБОТКА АРХИВНЫХ ДАННЫХ В СИСТЕМАХ ДИСПЕТЧЕРСКОГО УПРАВЛЕНИЯ В ЦЕЛЯХ ПОСТРОЕНИЯ ИНФОРМАЦИОННЫХ МОДЕЛЕЙ НЕШТАТНЫХ СИТУАЦИЙ

Л.И. Григорьев, Н.Е. Елов, И.В. Абдуллин

Описаны алгоритм поиска шаблонов и методика его применения для построения информационных образов нештатных и аварийных ситуаций в архивах событий систем управления технологическим процессом. Отмечено, что получаемые данные предназначены для использования в системах поддержки принятия диспетчерских решений, а также аналитических системах, компьютерных тренажерных комплексах подготовки диспетчерского персонала.

Ключевые слова: нештатные и аварийные ситуации, автоматизированная система диспетчерского управления, ассоциативные правила, оперативное управление, система поддержки принятия диспетчерских решений, SCADA-система.

ВВЕДЕНИЕ

Активное применение в АСУТП и, в том числе, в автоматизированных системах диспетчерского управления (АСДУ) SCADA-систем привело к появлению больших объемов архивной информации, которая, как показывает опыт, практически не используется. Благодаря внедрению современных средств автоматики и вычислительной техники снизилось число отказов функционирования технологических процессов (ТП), но при этом цена каждой аварии, связанной с нарушением технологического режима, существенно возросла. В этих условиях актуально создание систем поддержки принятия диспетчерских решений в нештатных и аварийных ситуациях [1].

Предлагается на основе анализа архивных данных в SCADA-системах сформировать информационные образы нештатных и аварийных ситуаций. Это позволит с учетом опыта диспетчеров построить модели нештатных и аварийных ситуаций и реализовать их в виде тренировочных упражнений на компьютерном тренажере диспетчерского управления. После всестороннего анализа решения таких задач на тренажере и создания со-

ответствующих алгоритмов управления разработанные алгоритмы в форме правил базы знаний следует включить в систему поддержки принятия диспетчерских решений (СППДР) в режиме реального времени.

Отправной точкой для решения указанной проблемы служит алгоритмизация методов обработки архивных данных.

1. СУЩЕСТВУЮЩИЕ СРЕДСТВА АНАЛИЗА АРХИВНОЙ ИНФОРМАЦИИ

При рассмотрении аварийной или нештатной ситуации можно выделить события, которые ее инициируют, а также события, непосредственно ей сопутствующие. В совокупности они составляют информационный образ ситуации, под которым понимается сценарий наступления и развития ситуаций, выходящих за рамки штатного функционирования ТП, и выраженный в определенной последовательности событий. Анализ архивной информации позволяет построить не только информационный образ нештатной или аварийной ситуации, но и спрогнозировать эти ситуации в реальном времени.

Ряд функций прогнозирования уже сейчас реализуется производителями комплексных систем автоматизации, однако основное распространение получили модули экономического планирования [2]. Предлагаемые средства анализа архивной информации («PI System/iFixHistorian», «Siemens WinCC ConnectivityPack», «InTouch SPCPro») обеспечивают потребности в формировании разного рода отчетной документации по ТП и полезны инженерам-технологам. Однако перечисленные средства существенно ограничены в гибкости настройки и взаимодействия с внешними информационными системами, что затрудняет их успешное применение в решении задач интеллектуального анализа данных [3]. Существуют также разработки, позволяющие на основе SCADA-систем реализовать функции СППДР. Чаще всего основные усилия разработчиков направлены на интеграцию СППДР и SCADA, а вопросам поиска знаний и формирования базы правил на основе архивных данных не уделяется достаточного внимания [4].

В этих условиях целесообразно создание универсального программного средства, осуществляющего поиск информационных образов в архивных данных SCADA-систем. Актуальность решения поставленной задачи подчеркивается широким применением SCADA-систем на самых разнообразных объектах и, следовательно, наличием значительных массивов «сырых» архивных данных, содержащих информацию о возможных режимах и ситуациях.

2. ОСОБЕННОСТИ РАБОТЫ С АРХИВАМИ SCADA-СИСТЕМ

Разработка алгоритмов обработки архивных данных SCADA-систем позволит создать необходимые информационные образы, а дальнейшее воспроизведение ситуаций (соответствующих найденным информационным образам) на комплексах моделирования технологических процессов или компьютерных тренажерах обеспечит диспетчерский персонал требуемыми знаниями для принятия решений в условиях, отличных от штатных.

В АСУТП существует надежный источник данных — архивы SCADA-систем, используемых в управлении. Основная проблема при использовании информации такого рода состоит в необходимости обработки достаточно больших массивов архивных данных в целях получения формализованных правил для применения в СППДР.

Исходные данные, находящиеся в архиве, должны быть преобразованы к форме, пригодной для применения алгоритма поиска информационных образов, и требуют выполнения ряда промежуточных этапов. Принимая во внимание большое раз-

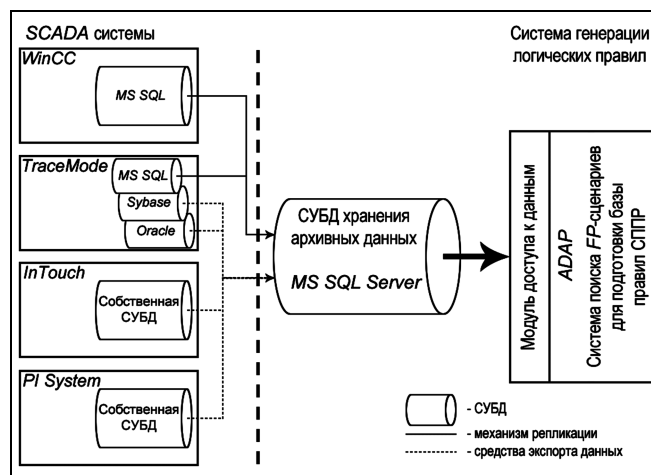


Рис. 1. Схема обмена информацией с АСУТП

нообразие SCADA-систем, логично предположить, что системы хранения данных также отличаются друг от друга. Использование для каждой системы хранения данных своего уникального модуля (с единым интерфейсом общения с модулем, реализующим алгоритм поиска информационных образов) позволяет унифицировать представление данных и обеспечить легкость масштабирования и адаптации системы к новым или изменившимся данным при расширении или модернизации системы управления. В связи с этим процедура обращения к архивной информации будет изменяться в различных системах без необходимости изменения логики работы основного алгоритма (рис. 1).

3. МЕХАНИЗМЫ ОБРАБОТКИ АРХИВНЫХ ДАННЫХ

Развитие нештатных или аварийных ситуаций представляет собой последовательность нескольких связанных друг с другом событий. В интеллектуальном анализе данных такого рода последовательности называются ассоциациями [3].

Для формирования ассоциативных правил на основе архивных данных необходимо обеспечить эффективный поиск информационных образов по всему объему истории событий. После ряда преобразований модель хранения данных в архивах систем управления становится пригодной для применения в алгоритмах поиска ассоциаций.

Далее представлена математическая формулировка ассоциативных правил, позволяющих находить закономерности между связанными событиями.

Пусть $I = \{i_1, i_2, i_3, \dots, i_n\}$ — набор элементов, а D — множество транзакций, где каждая транзак-



ция T — это набор элементов из I , $T \subseteq I$. Каждая транзакция представляет собой бинарный вектор, где $t[k] = 1$, если i_k элемент присутствует в транзакции, иначе $t[k] = 0$. Говорят, что транзакция T содержит X , некоторый набор элементов из I , если $X \subseteq T$. Ассоциативным правилом называется импликация $X \Rightarrow Y$, где $X \subseteq I$, $Y \subseteq I$ и $X \cap Y = \emptyset$.

В результате реализации алгоритма поиска ассоциативных правил сложно формализуемый архив технологических событий трансформируется в существенно более компактный вид, где весь список событий представляется в виде конечных наборов элементов, повторяющихся разное число раз в определенной последовательности. Такие наборы легко трансформируются в искомые правила, описывая историю развития различных технологических сценариев. На основе выработанных таким образом правил появляется возможность строить различные системы, работающие на основе правил в виде продукций и обеспечивающие мониторинг процесса в реальном времени в смысле приближения к тому или иному режиму функционирования системы, описанному в виде последовательности событий.

Организация хранения данных в архивах SCADA-систем плохо сочетается с применением алгоритмов анализа информации. Поэтому одна из главных задач на этапе получения данных из архива SCADA-системы состоит в преобразовании их к единому формату, содержащему, как минимум, идентификатор события и время его наступления.

4. ВЫБОР МЕТОДА ФОРМИРОВАНИЯ ИНФОРМАЦИОННЫХ ОБРАЗОВ

Исследование различных методов поиска ассоциативных правил выявило преимущества алгоритма FPG (Frequent-Pattern Growth Strategy), что также подтверждается данными компании «BaseGroupLabs», известного производителя информационных систем бизнес-аналитики. Например, в сравнении с алгоритмом «Apriori» обеспечивается существенно более высокое быстродействие при увеличении объемов обрабатываемых данных [5].

Кроме того, для алгоритма FPG процедура преобразования модели данных SCADA в пригодный для анализа вид выглядит наиболее простой. Изначально методы поиска ассоциаций были направлены на работу с данными экономического характера, однако применение их в сфере анализа архивных данных технологических систем не встречает дополнительных трудностей, кроме необходимости первичной подготовки информации. Требования

секвенциального анализа, т. е. обеспечение не только поиска схожих по составу информационных портретов, но и проверки строгого соблюдения временной последовательности событий во всех шаблонах, вносят дополнительную сложность. Определенный интерес могут представлять и шаблоны, не удовлетворяющие требованиям к строгой последовательности событий, поэтому их поиск и фиксация также реализуются в рамках алгоритма анализа.

Для определения информационного множества событий в описании алгоритма генерации частых шаблонов вводится понятие транзакции. При рассмотрении модели данных архивов SCADA главной задачей становится определение критерия связи между последовательными событиями, поскольку основной интерес представляет изменение состояния объекта в аварийных и нештатных ситуациях. Предполагается, что в каждый момент времени может произойти весьма мало событий, и такой временной срез не отражает в полной мере схему развития аварии.

5. АЛГОРИТМ ПОИСКА ИНФОРМАЦИОННЫХ ОБРАЗОВ И ЕГО ПРОГРАММНАЯ РЕАЛИЗАЦИЯ

Программная реализация алгоритма анализа архивных данных состоит из этапов:

- 1) получения данных из архива SCADA-системы;
- 2) преобразования полученных данных в удобный для программного анализа формат;
- 3) выделения серий событий по временному признаку;
- 4) выделения частых шаблонов.

На первом этапе работает механизм получения информации из БД используемой SCADA-системы и сохранения этих данных в едином внутреннем формате для обеспечения простоты масштабирования и обновления системы. Обзор наиболее распространенных программных решений АСУТП показал, что для организации хранения архивных данных большинство производителей либо используют в качестве базовой СУБД, либо предлагают опционально «Microsoft SQL Server».

При выборе механизма доступа к данным технология ADO, как более современная и производительная, оказалась предпочтительнее по сравнению с технологией ODBC. Она обеспечивает поддержку большинства современных СУБД при необходимости интеграции с отличными от «Microsoft SQL Server» программными платформами (рис. 1).

На втором этапе архивы SCADA преобразуются к модели данных в системе анализа, с учетом выполнения двух противоречивых требований: сокращения требуемых вычислительных ресурсов и обеспечения максимальной производительности

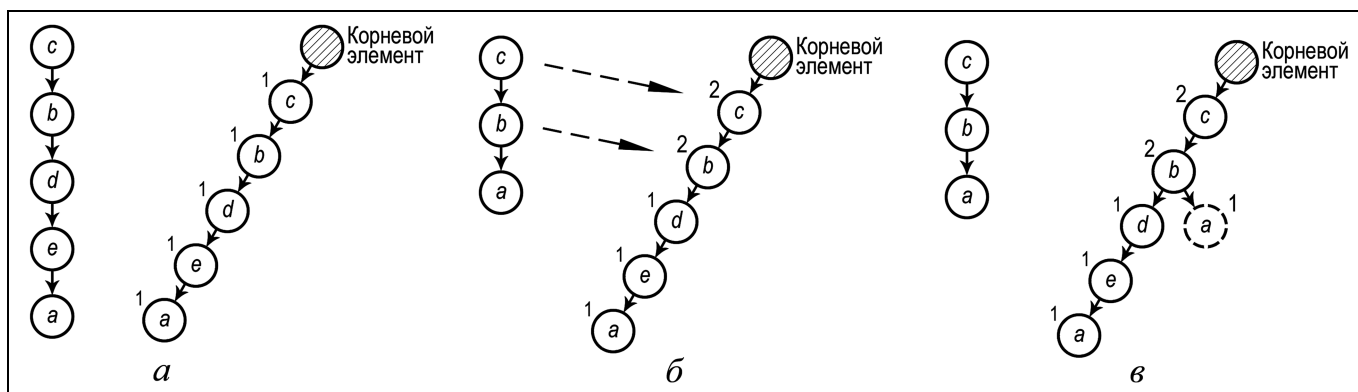


Рис. 2. Алгоритм FPG: построение дерева

вычислений. Оба этих ограничения связаны с тем, что архивные данные, накопленные за несколько лет, могут достигать больших объемов. Следовательно, необходимо обеспечить приемлемые системные требования по объему оперативной памяти и предоставить наиболее полный набор данных во избежание необходимости повторных подключений к серверу СУБД. Учитывая это и возможное многообразие форм представления событий в БД различных SCADA-систем, предлагается ограничиться рассмотрением идентификатора события, его статусом, а также временными характеристиками, такими как время наступления события (год/месяц/день часы:минуты:секунды) и уточняющего его параметра (в миллисекундах). Итак, для первичного анализа данных в качестве основных служат временные характеристики.

На третьем этапе работы алгоритма архив независимых дискретных событий преобразуется в серии, формируемые на основе критерия временной близости (временной коэффициент)

$$K_t = \frac{C}{N-1} \sum_{i=1}^{N-1} \Delta t_i, \text{ где } \Delta t_i - \text{разница меток времени}$$

событий i и $i+1$, N — общее число событий, C — определяемый экспериментально поправочный коэффициент. Использование вычисляемого временного коэффициента в качестве базового обеспечивает автоматическое выполнение процедуры разделения последовательности событий, а в ручном режиме изменением его значения достигается необходимая гибкость вычисления: $i_n \in T$, если $i_{n-1} \in T$ и $t_{i_n} - t_{i_{n-1}} \leq K_p$, где K_t — выбранный временной коэффициент.

Полученная в итоге преобразований структура хранения данных применяется в реализуемом на четвертом этапе алгоритме генерации информационных образов. По завершению поиска наборов FP (Frequent-Pattern — «часто встречающиеся

предметные наборы») возможна следующая итерация — определение для каждого из найденных FP-наборов набора-предшественника. Эта процедура на большем временном отрезке позволяет выделить долговременные причинно-следственные связи и служит для обеспечения заблаговременной классификации текущей ситуации при помощи СППДР.

Алгоритм FPG реализован в соответствии с его распространенным описанием [6–8] и учитывает специфику модели хранения технологических архивных данных.

Алгоритм состоит из двух основных процедур: формирования дерева частых предметных наборов и процедуры выбора наборов, удовлетворяющих заданным условиям из созданного дерева. При построении дерева действует следующее правило: «Если для очередного события в дереве встречается узел, идентификатор которого совпадает с идентификатором события в серии, то программа не создает нового узла, а увеличивает вес соответствующего узла в дереве на 1. В противном случае для этого события создается новый узел и ему присваивается вес 1». Для удобства предоставления и наглядности информации события представлены буквенными идентификаторами (рис. 2).

Как видно на рис. 2, а, для каждого события из серии «с—b—d—e—a» будет создан новый узел и, таким образом, первая ветвь дерева. Для событий с и b следующей серии с—b—a, новые узлы созданы не будут, но вес уже существующих соответствующих узлов увеличится на 1 (рис. 2, б). Для события a из той же серии будет создан новый узел, так как событие a не совпадает с последующим узлом d (рис. 2, в).

Итоговые наборы формируются от ветвей к корню дерева (рис. 3). Каждый частый предметный набор состоит из суффикса, префикса и подержки, причем префикс — это путь в дереве до

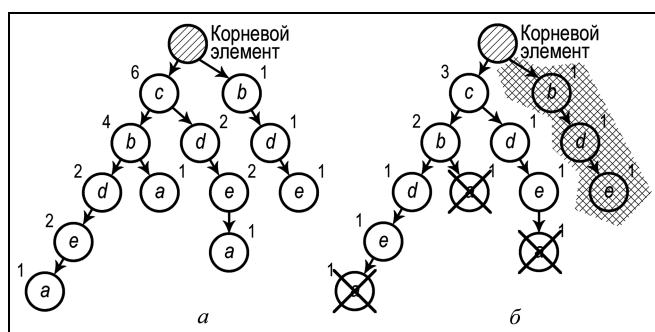


Рис. 3. Алгоритм FPG: извлечение наборов

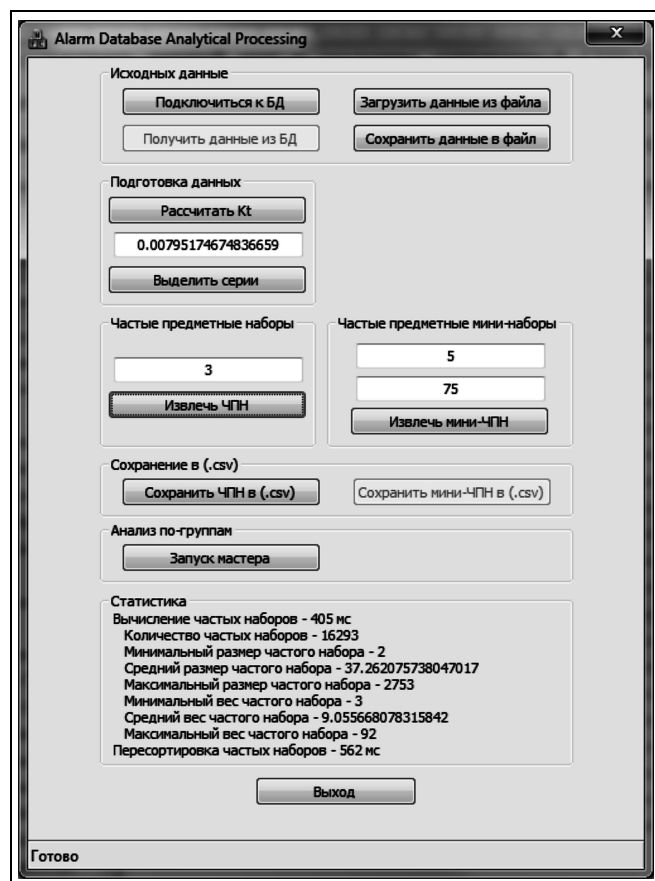


Рис. 4. Пользовательский интерфейс

элемента, суффикс — сам элемент, а поддержка — число аналогичных путей в дереве. Именно поддержка служит эффективным средством классификации отобранных наборов для описания штатных, нештатных и аварийных режимов. На рис. 3, а продемонстрировано сформированное и подготовленное к процедуре выборки дерево, указаны веса элементов; на рис. 3, б изображен поиск наборов для суффикса *a*, исключены ветви, не содержащие искомого суффикса.

Для управления процедурой анализа предусмотрена возможность задания ряда параметров, влияющих на точность, полноту и скорость работы алгоритма. Помимо временного коэффициента, этими параметрами являются минимальный вес для FP-набора, временной период и относительный вес для наборов, предшествующих найденным.

Текущая версия программного продукта представляет собой независимо функционирующую среду исполнения с собственным пользовательским интерфейсом (рис. 4) и предоставляет ряд диагностических и информационных сообщений, позволяющих оценить эффективность аналитического процесса и при необходимости корректировать входные параметры. Доступным пользователю результатом обработки является экспортированный с заданными параметрами в формате Excel перечень вычисленных FP-наборов, наборов, предшествующих FP-наборам, и всех определяемых в процессе анализа параметров.

6. ПЕРСПЕКТИВЫ КЛАССИФИКАЦИИ НАЙДЕННЫХ ШАБЛОНОВ И ОТЛАДКА ПРОГРАММНОГО ПРОДУКТА

Полученные данные хорошо приспособлены для оценки и визуализации средствами Excel. Для интеграции с СППДР предусмотрена возможность межпроцессного взаимодействия как в рамках локальной ЭВМ, так и посредством локальной вычислительной сети.

Для описания принципов работы с данными предлагается следующая двухуровневая модель. Пусть имеется набор данных уровня процесса, т. е. каждый информационный элемент описан по месту инициализации. Такое контекстное представление данных соответствует требованиям системы управления процессом и выполняет одновременно две функции: обеспечивает идентификацию данных в архиве и предоставляет пользователю систему адресации, характерную для информационного описания объекта управления. В случае использования архивных данных в аналитических алгоритмах, необходимость такого рода контекстной адресации отпадает, кроме того, излишняя информационная насыщенность усложняет сам процесс анализа и ведет к увеличению требований к объему оперативной памяти. Однако при интерпретации результатов необходимо предусмотреть возможность оперативного получения необходимой иерархической информации из БД. Уровень данных системы анализа можно охарактеризовать как абстрактный или контекстно-независимый (рис. 5).

В большинстве случаев в проектах АСДУ на основе SCADA-системы создается классификация

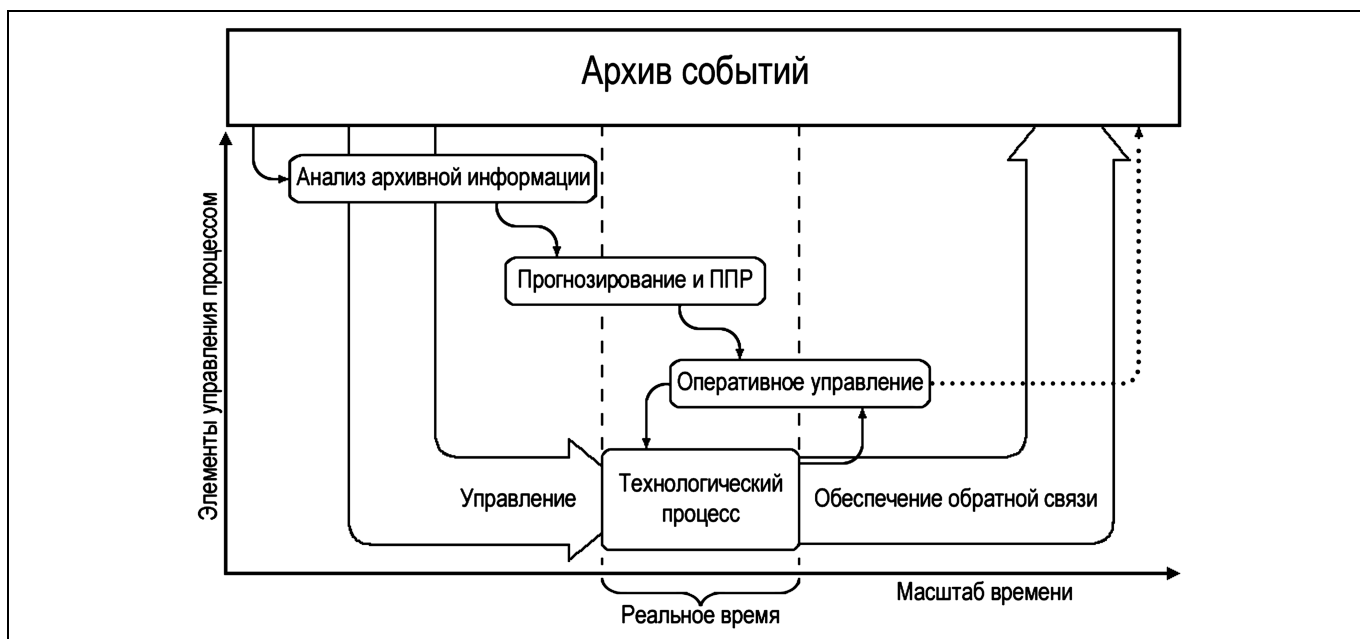


Рис. 5. Схема обратной связи по ретроспективной информации

сообщений по степени их опасности (аварийные, предупредительные, информационные, диагностические и др.). В рамках реализации алгоритма анализа архивных данных предлагается технология классификации найденных FP-наборов на основе данных о ранжировании в системе сообщений базовой SCADA-системы. Классификатор позволяет создавать динамические классы данных, указывая принадлежность к ним на базе классов SCADA, и затем определять для каждого FP-набора относительные доли по каждому из определенных пользователем классов. Такой механизм позволяет сортировать весь объем вычисленных наборов по важности и вероятности возникновения аварии. Взаимосвязь между двумя уровнями представления данных обеспечивает необходимую привязку к рассматриваемому объекту, а само разбиение позволяет более рационально организовать вычислительный процесс обработки архивных данных.

Для отладки программного обеспечения использовалась архивная информация из БД реального объекта — электростанции на попутном газе для нужд нефтедобывающего участка. Система SCADA реализована на базе продукции фирмы «Siemens» — «SIMATIC WinCC» с применением подсистемы «SICAM PAS» и организацией связи со сторонними контроллерами посредством OPC протокола. Система верхнего уровня по объему информации охватывает все ТП на электростанции, что обеспечивает достаточную информативность результатов анализа. Архив накоплен при-

мерно за полтора года и содержит около 150 тыс. информационных записей. Программный модуль реализован с учетом конфигурации данного объекта для работы с сервером СУБД «Microsoft SQL», который поддерживается технологией доступа к данным ADO.

На базе тестового архива исследовалась эффективность работы программной реализации алгоритма, и выполнялись отладка и оптимизация программного кода. В результате удалось обеспечить требуемое для расчетов быстродействие, обеспечивающее поиск правил по мере появления новых данных.

ЗАКЛЮЧЕНИЕ

Методика подготовки и анализа архивных данных SCADA-систем позволяет осуществить первый шаг к созданию современной СППДР в области управления ТП, отвечающей жестким требованиям к скорости работы в режиме реального времени. Применение механизмов такого рода в условиях постоянного роста стоимости ошибки — важное направление повышения надежности человеко-машинных систем управления.

Предлагаемый алгоритм позволяет перейти к качественно новому уровню оценки возникающих на объекте аварийных и нештатных ситуаций. На основе статистически выявленных закономерностей появляется возможность введения подробной



и автоматизированной классификации ситуаций. Вычисление ассоциативных правил затрагивает различные по временному масштабу ретроспективные выборки, что позволяет выявить не только шаблоны аварийных ситуаций, но также и достаточно развернутые первопричины их возникновения. Более того, при наличии достаточно больших архивов аналитический инструментарий позволяет вычислять динамику изменения аварийности при различных классификационных условиях, что может использоваться как для формирования различных отчетных данных, так и для интеграции с информационными системами верхнего уровня при решении задач контроля качества и обеспечения надежности системы автоматизированного диспетчерского управления.

ЛИТЕРАТУРА

1. Григорьев Л.И. Автоматизированное диспетчерское управление — магистральное направление развития АСУТП газовой отрасли // Газовая промышленность. — 2010. — № 3. — С. 76—83.
2. Башлыков А.А. Принципы построения средств интеллектуальной поддержки принятия решений диспетчером ТДП в ЕСУ ТС ВСТО // Автоматизация, телемеханизация и связь в нефтяной промышленности. — 2009. — № 11.
3. Дюк В., Самойленко А. Data Mining: учебный курс — СПб.: Питер, 2001. — 368 с.
4. Концепция развития функций SCADA-системы TRACE MODE на основе технологии экспертных систем принятия и исполнения решений / А.С. Береза, В.П. Прохоров, А.В. Прохоров, А.Е. Крохмаль // Информатизация и системы управления в промышленности. — 2005. — № 1 (5).
5. Орешков В. FPG — альтернативный алгоритм поиска ассоциативных правил. — URL: http://www.basegroup.ru/library/analysis/association_rules/fpg/ (дата обращения 16.11.2011).
6. Agrawal R., Imielinski T., and Swami A. Mining association rules between sets of items in large databases // In Proc. of the ACM SIGMOD Conference on Management of Data, Washington, D. C., May 26—28. — 1993.
7. Agrawal R., Srikant R. Fast Discovery of Association Rules // In Proc. of the 20th Intern. Conf. on VLDB, Santiago, Chile, 1994.
8. Han J., Pei J., Yin Y., Mao R. Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach // Data Mining and Knowledge Discovery. — 2004. — N 8.

Статья представлена к публикации членом редколлегии В.Г. Лебедевым.

Григорьев Леонид Иванович — д-р техн. наук, профессор, зав. кафедрой, Российский государственный университет нефти и газа им. И.М. Губкина, г. Москва,
☎ (499) 135-71-56, ✉ lgrig@gubkin.ru,

Елов Николай Евгеньевич — гл. инженер проектов, ООО «ЭнергопромАвтоматизация», г. Москва,
☎ (499) 235-21-84, ✉ elov@epsa-spb.ru,

Абдуллин Игорь Валерьевич — студент, Российский государственный университет нефти и газа им. И.М. Губкина, г. Москва,
☎ (499) 125-44-98, ✉ pcmankiller@mail.ru.

Новая книга

Трахтенгерц Э.А. Компьютерные технологии манипулирования общественным мнением. — М.: СИНТЕГ, 2011. — 296 с.

Рассмотрены особенности информационного общества и компьютерного мониторинга состояния его отдельных групп и массовых аудиторий. Показаны компьютерные алгоритмы формирования целей, стратегий и оперативных действий манипулирования общественным мнением. Обсуждены компьютерные методы генерации оценок, позволяющих манипулировать общественным мнением, а также коррекции целей, стратегических решений и оперативных действий в динамике манипулирования общественным мнением.

Для широкого круга читателей.