

на правах рукописи

СУРПИН ВАДИМ ПАВЛОВИЧ

ПРЕДМЕТНО-НЕЗАВИСИМЫЕ МОДЕЛИ МНОГОКОМПОНЕНТНЫХ
СИСТЕМ И ИХ ПРИМЕНЕНИЕ В СИСТЕМАХ МОНИТОРИНГА

05.13.01 – Системный анализ, управление и обработка информации
(технические системы)

автореферат диссертации на соискание учёной степени
кандидата технических наук

Москва 2012

Работа выполнена в Федеральном государственном бюджетном учреждении науки Институте проблем передачи информации им. А.А. Харкевича Российской академии наук

Научный руководитель: кандидат технических наук, доцент
Баканова Нина Борисовна

Официальные оппоненты: Бернштейн Александр Владимирович,
доктор физико-математических наук, профессор,
ИСА РАН, заведующий лабораторией

Степановская Ираида Александровна,
кандидат технических наук, старший научный
сотрудник,
ИПУ РАН, ведущий научный сотрудник

Ведущая организация: Федеральное государственное бюджетное учреждение науки Институт проблем информатики Российской академии наук

Защита состоится «__» _____ 20__ г. в _____ на заседании диссертационного совета Д 002.077.01 при Институте проблем передачи информации им. А.А. Харкевича РАН (127994, г. Москва, ГСП-4, Большой Каретный переулок, 19, стр.1.).

С диссертацией можно ознакомиться в библиотеке ИППИ РАН.

Автореферат разослан «__» _____ 2012 г.

Учёный секретарь
диссертационного совета
доктор физико-математических наук

И.И. Цитович

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность исследования. Мониторинг явлений, процессов и их характеристик необходим для поддержки научно обоснованного принятия решений в процессе функционирования системы. В широком смысле, под мониторингом понимают систематический сбор и анализ информации об объекте с целью идентификации отклонений его характеристик от штатного режима, либо выявления тенденции изменения характеристик.

Задача мониторинга характерна для тех видов деятельности, где изучаемый объект представляет сложную многокомпонентную систему, о состоянии которой можно судить по изменению свойств системы в целом или совокупности составляющих её объектов. В качестве примеров видов деятельности, неотъемлемой частью которых является задача мониторинга, можно привести эпидемиологический надзор, управление крупномасштабными техническими объектами и производствами, организационное управление.

Многочисленные исследования в каждой из указанных областей, а также во многих других, указывают мониторинг как инструмент исследования сложных систем, так и практически значимую задачу, без которой невозможно нормальное функционирование многих организационных систем и технологических процессов, а также управление ими. Методы мониторинга традиционно применяются в экологических исследованиях. Проблемам экологического мониторинга посвящены работы И. П. Герасимова, Ю.А. Израэля, В.А. Королева, В.Д. Минченко. Исследованию методов мониторинга в здравоохранении посвящены работы Э. Лавсона, К. Клейнмана, Дж. Ломбардо, Д. Бакериджа, В. И. Покровского, Н.Н. Филатова, И. Л. Шананина. Вопросам технической диагностики посвящены работы А.В. Мозгалевского, П.П. Пархоменко, Е.С. Согомоняна, И.А. Биргера, Я.А. Фомина, Л.П. Глазунова, Б.В. Павлова, В.В. Ключева.

Несмотря на широкое применение систем мониторинга во многих областях деятельности, в существующих исследованиях отсутствует достаточно общая модель этих систем, которая бы позволила бы систематизировать накопленный опыт и эффективно использовать его. В каждой области создание системы мониторинга начинается «с нуля», использование наработок из смежных областей затруднено, а передача знаний возможна лишь с опытом разработчика. Это негативно сказывается на общей стоимости, сроках разработки и качестве проектируемых систем. В связи с этим, задача создания обобщённой модели систем мониторинга является актуальной. Препятствием к созданию обобщённой модели является значительная зависимость алгоритмического и математического обеспечения мониторинга от предметной области, поэтому для решения проблемы требуется решить задачу изоляции описания предметной области от общих методов и алгоритмов. Предметная область может быть описана информационными моделями, применяемыми при построении экспертных систем. В работе предлагается использовать информационные модели, используемые в экспертных системах, при проектировании систем мониторинга.

Таким образом, **цель исследований** состоит в разработке обобщённой предметно-независимой модели систем мониторинга и методов построения систем с использованием предложенной модели. Для достижения цели исследования поставлены и решены следующие **задачи**:

- Разработка формальной предметно-независимой модели многокомпонентной системы, являющейся предметом мониторинга, и динамики изменения её состояния;
- Анализ существующих информационных моделей представления знаний и обоснованный выбор информационной модели представления знаний предметной области для использования в интеллектуальных алгоритмах систем мониторинга;
- Разработка методов интеллектуального анализа и обработки экспертной информации для использования в предметно-независимых системах мониторинга многокомпонентных систем;

Методы исследований. В работе использованы методы общей теории систем, теории управления, методов математического анализа данных, теории графов, теории исследования операций, а также методов объектно-ориентированного анализа и проектирования систем.

Научная новизна заключается в комплексном исследовании систем мониторинга, в составе которых используются экспертные оценки состояния объектов мониторинга:

- Предложена формальная теоретико-множественная модель многокомпонентной системы мониторинга, а также объектная модель для проектирования информационных систем;
- Проведён анализ существующих формальных моделей представления знаний предметной области. На основании проведённого анализа выбрана информационная модель представления знаний предметной области для использования в интеллектуальных алгоритмах систем мониторинга;
- Разработаны методы интеллектуального анализа и обработки экспертной информации с использованием информационной модели предметной области, в том числе способ оценки неопределённости характеристик объектов мониторинга, а также алгоритм кластеризации объектов для формирования гипотезы о характере внешнего воздействия.

Практическая значимость исследования и реализация результатов. Научные результаты, полученные в диссертации, использованы при создании программно-аппаратного комплекса информационной системы мониторинга распространения вирусного гепатита, а также при создании системы планирования и мониторинга контрольных мероприятий Федеральной службы по надзору в сфере транспорта. Результаты использованы в рамках выполнения следующих работ:

- Федеральная целевая программа «Предупреждение и борьба с социально значимыми заболеваниями (2007-2011 годы)», проект номер 29-10/07 «Разработка концепции и моделирующего комплекса эпидемиологии острых и хронических вирусных гепатитов для нужд ФГУН "ЦНИИЭ" Роспотребнадзора»;
- Федеральная целевая программа «Предупреждение и борьба с социально значимыми заболеваниями (2007-2011 годы)», проект номер ИН-0008/08 «Разработка проектных решений и программных модулей графического представления эпидемиологических данных»
- Ведомственная целевая программа Федеральной службы по надзору в сфере транспорта «Повышение качества исполнения государственных функций и предоставления государственных услуг, в том числе, в электронном виде», ГК №44 «Разработка информационной системы контрольных мероприятий и документооборота в Ространснадзоре»

Достоверность научных положений определяется корректным использованием математического аппарата и практической реализацией конкретных информационных систем.

На защиту выносятся:

1. Предметно-независимые модели мониторинга многокомпонентных систем: формальная теоретико-множественная модель многокомпонентной системы мониторинга, а также объектная модель для проектирования информационных систем, описывающие структуру многокомпонентных систем мониторинга, для которых применимы предложенные в работе методы.
2. Анализ существующих формальных моделей представления знаний предметной области и обоснованный выбор информационной модели представления знаний предметной области, позволяющую создать базу знаний, выполняющую необходимые для использования в интеллектуальных алгоритмах систем мониторинга запросы;
3. Методы интеллектуального анализа и обработки экспертной информации с использованием информационной модели предметной области, включая способ оценки неопределённости характеристик объектов мониторинга и алгоритм кластеризации объектов мониторинга для формирования гипотезы о характере внешнего воздействия, позволяющие идентифицировать недостоверные данные мониторинга;
4. Прототип программного комплекса, созданный на основе предложенных моделей и методов, осуществляющий сбор экспертных характеристик объектов мониторинга, и их интеллектуальную обработку с использованием проблемно-ориентированных семантических графов для обеспечения информационной поддержки при принятии управленческих решений, который демонстрирует применимость разработки для различных предметных областей.

Апробация работы. Основные положения и результаты диссертационной работы были доложены и обсуждались на всероссийской конференции «XII Всероссийский симпозиум по прикладной и промышленной математике (весенняя сессия)», международной конференции «Управление развитием крупномасштабных систем (MLSD'2009)», всероссийской конференции «III Всероссийская молодёжная конференция по проблемам управления (ВМКПУ'2008)», конференции молодых учёных и специалистов «Информационные технологии и системы (ИТиС'2008)», международном семинаре «Распределённые компьютерные и телекоммуникационные сети (DCCN'2005)».

Публикации. Основное содержание работы отражено в 10 работах, в том числе в журнале «Информационные технологии», рекомендованном ВАК РФ для изложения основных научных результатов диссертации на соискание учёных степеней.

Структура и объём работы. Диссертация состоит из введения, четырёх глав, приложений и списка литературы (69 наименований). Имеет общий объём 153 машинописных страницы, содержит 44 рисунка и 8 таблиц.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** обосновывается актуальность темы диссертации, формулируются общие проблемы, цели и задачи исследования, научное и практическое значение полученных результатов, рассматривается структура диссертации и взаимосвязь отдельных глав.

В **первой главе** рассмотрены основные требования к системам мониторинга; представлен обзор математических методов, нашедших применение при решении задач мониторинга систем и процессов; приводятся основные сведения о построении информационных моделей предметных областей и существующие подходы к формализации знаний. Основное содержание главы опубликовано в [1, 7, 8].

Во **второй главе** предложена формальная теоретико-множественная модель многокомпонентной системы мониторинга, поставлены формальные задачи системы мониторинга в предложенной модели. Основное содержание главы опубликовано в [9, 10].

В **третьей главе** построена объектная модель системы мониторинга. На основе предложенной модели определены принципиальные свойства взаимодействующих объектов, предложен шаблон описания сложной системы, позволяющий формализовать характерные для системы задачи и применить к ним предложенные методы и алгоритмы. Основное содержание главы опубликовано в [4, 5].

В **четвёртой главе** приводятся результаты применения разработанных моделей и методов при разработке конкретных информационных систем – описано приложение предложенных методов и алгоритмов к проектированию и разработке

систем мониторинга для эпидемиологии и транспортной безопасности. Основное содержание главы опубликовано в [1, 2, 3].

В первой главе:

- Рассмотрены основные требования к системам мониторинга;
- Приводится обзор математических методов, положенных в основу методов мониторинга;
- Приводятся сведения из области экспертных систем о средствах представления знаний предметной области, а также о вероятностных сетях представления причинно-следственных связей.

В **разделе 1.1** обозначен ряд качественных характеристик, отражающих общие требования к сервисам мониторинга, среди которых сложность, гибкость, качество данных, приживаемость, чувствительность, прогностичность положительного результата, представительность, оперативность, стабильность. Рассмотрено влияние отдельных характеристик на выбор подходов к созданию систем мониторинга.

В **разделе 1.2** приводится обзор математических методов обнаружения разладки целевой функции, нашедших применение в мониторинге систем и процессов: метод наименьших квадратов, авторегрессионное интегрированное скользящее среднее, экспоненциально взвешенное скользящее среднее, обобщённая линейная модель, метод накопленных сумм, вейвлет-анализ, метод опорных векторов, скрытая Марковская модель, методы на основе Байесовских сетей. Приведены краткие описания основных идей, заложенных в перечисленные математические методы.

В **разделе 1.3** приведены сведения из области экспертных систем о средствах представления знаний предметной области, а также о вероятностных сетях представления причинно-следственных связей. Приведены описания типовых структур представления данных структур знаний. Отдельно дан обзор вероятностных сетей представления знаний, способов обучения интеллектуальных структур на их основе.

Во **второй главе** предложена формальная теоретико-множественная модель многокомпонентной системы мониторинга, поставлены формальные задачи системы мониторинга в предложенной модели.

- Описана общая структура модели, определены характеристики составляющих её элементов (микроуровень) и наблюдаемые характеристики системы как единого целого (макроуровень);
- Предложена модель изменения состояния объектов на микроуровне, приведён пример модели изменения состояний на микроуровне объектов для эпидемиологии;

- Предложена модель описания основных характеристик на макроуровне, приведён пример модели изменения состояний характеристик системы объектов мониторинга для эпидемиологии;
- Формализованы задачи мониторинга в терминах предложенной теоретико-множественной модели.

В разделе 2.1 описана общая структура модели. Исследуемая система состоит из множества объектов мониторинга, взаимодействующих между собой, и множества источников информации, измеряющих параметры объектов и передающие их в центр обработки данных (рис. 1).

На схеме введены следующие обозначения: $\{W, I\}$ - многокомпонентная система, $M = \{S, E\}$ - система мониторинга. Многокомпонентная система $\{W, I\}$ представлена множеством $W = \{w_1, \dots, w_{|W|}\}$ объектов мониторинга и моделью I изменения состояния объектов. Объект $w \in W$ описывается вектором свойств $D = (d_1 \dots d_{|D|})$ размерности $|D|$, $d_i \in \mathbb{R}, i = 1 \dots D$. Система мониторинга $M = \{S, E\}$ состоит из множества источников информации $S = \{s_1, \dots, s_{|S|}\}$ и центра обработки данных E .

Для моделирования динамики системы в разделе 2.2 работы использован подход к моделированию сложных систем, характерной чертой которого является формулирование гипотез, описывающих динамику процессов на микроуровне, ведущих к пониманию динамики основных агрегированных характеристик на макроуровне. Модель I описывает динамику изменения состояния D отдельных объектов $w \in W$ во времени и может быть представлена в следующем виде. Состояние объекта W описывается вектором D , и изменяется в некоторые дискретные моменты в соответствии с правилом:

$$D_{\text{нов.}} = G(D, A(t), F(t)),$$

где $D_{\text{нов.}}$ - новое состояние объекта, D - предыдущее состояние объекта, $A(t)$ - матрица взаимодействия объектов мониторинга, $F(t)$ - вектор внешнего воздействия на систему объектов мониторинга. Вектор внешнего воздействия и матрица взаимодействия объектов зависят от времени.

Частным случаем модели является линейная модель, которая может описать основные свойства поведения реальных систем или использоваться как первое приближение. Уравнение описывает изменение состояние объекта, обусловленное внутренними законами его функционирования, внешним воздействием f и

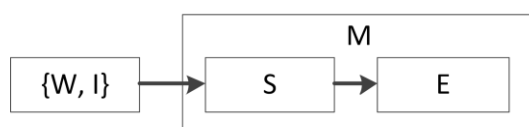


Рис. 1. Общая схема исследуемой системы

взаимодействием между объектами $w \in W$:

$$\mathbf{D}_i(T_{n+1}) = \mathbf{a}_{ii} \cdot \mathbf{D}_i(T_n) + \sum_{j=1, i \neq j}^{|W|} \mathbf{a}_{ij} \cdot \mathbf{D}_j(T_n) + \mathbf{f}_i(T_n),$$

где $\dots < T_0 < T_1 < \dots < T_n < \dots$ - моменты изменения состояния объекта W , член $\mathbf{a}_{ii} \cdot \mathbf{D}_i(T_n)$ описывает внутренние законы изменения состояния объекта, член $\sum_{j=1, i \neq j}^{|W|} \mathbf{a}_{ij} \cdot \mathbf{D}_j(T_n)$ описывает взаимодействие объектов, а $\mathbf{f}_i(T_n)$ описывает воздействие внешней среды на объект w_i .

В заключение раздела приводится пример применения предложенной модели к описанию эпидемиологического процесса передачи инфекции.

В разделе 2.3 предложена модель описания основных характеристик на макроуровне. С точки зрения мониторинга систем и процессов интересны агрегированные характеристики системы, отражающие динамику процессов на макроуровне. Эти показатели, такие как уровень заболеваемости, преступности и т.д., часто отражают зависимость величин $|C|$ подмножеств C множества объектов W от времени. Подмножества вводятся по принципу близости вектора состояния \mathbf{D} объекта к центральному вектору подмножества:

$$w \in C \Leftrightarrow \rho_C(\mathbf{D}, \mathbf{D}_C) < \varepsilon_C,$$

где w – объект из множества объектов W , $C \subset W$ – подмножество множества объектов W , ρ_C – расстояние между векторами состояний объектов, \mathbf{D} – вектор состояния объекта w , \mathbf{D}_C – центральный вектор подмножества C , ε_C – радиус подмножества C . Таким образом, подмножество C задаётся тремя параметрами: $C = C(\rho_C, \mathbf{D}_C, \varepsilon_C)$.

Каждому объекту w может соответствовать одно или более подмножеств, в этом случае будем говорить, что объект принадлежит одному или более классам: $C(w) = C(\mathbf{D}) = \{C_{w1}, \dots\}$, где \mathbf{D} – вектор состояния объекта w . Поскольку $\mathbf{D} = \mathbf{D}(t)$, то и набор классов объекта меняется во времени, то есть $C(w) = C(t)$.

Величина $|C|$ класса C определяется как количество объектов w , принадлежащих классу C : $|C| = \sum_{w \in W} w \in C$, где выражение $w \in C$ принимает значение 1, если объект w принадлежит классу C и 0 в противном случае. В терминах объектов и классов агрегированные показатели, характеризующие динамику процессов на макроуровне, принимают вид

$$F(t) = F(|C_1(t)|, |C_2(t)|, \dots, |C_n(t)|, t),$$

где $F(t)$ – агрегированный показатель состояния системы, C_1, \dots, C_n – классы объектов, t – время. В данном выражении функциональная зависимость F от величин классов $|C|$ обозначает как зависимость от самой функции $|C(t)|$, так и от её производных, интегральных и других возможных характеристик.

В заключение раздела приводится пример применения предложенной модели к описанию макро-характеристик эпидемиологического процесса.

В разделе 2.4 сформулированы задачи мониторинга состояния системы W в терминах предложенной математической модели. Рассмотрены две задачи: задача идентификации отклонения характеристик системы от штатного режима и задача проверки гипотез о характере отклонения.

Первая заключается в выборе набора классов C_1, \dots, C_n , функции $F(t)$, методов её измерения, штатного режима $F_0(t)$ и порога отклонения ε_F таких, что отклонение $|F(t) - F_0(t)| > \varepsilon_F$ будет соответствовать существенной коррекции модели изменения состояний объектов I . Понятие существенного отклонения зависит от конкретной области применения модели и может быть формализована, например, в терминах устойчивости системы W .

Вторая задача расширяет первую. Помимо идентификации отклонения $|F(t) - F_0(t)| > \varepsilon_H$ необходимо выбрать наиболее вероятную гипотезу $H_{prob}(t)$ из множества гипотез $H = \{H_1, \dots, H_N\}$ такую, что $|F(t) - H_{prob}(t)| = \min_{H_i \in H} (|F(t) - H_i(t)|)$.

Выигрыш от проверки гипотез развития процесса заключается в выборе значения $\varepsilon_H < \varepsilon_F$ и, как следствие, раннем обнаружении отклонения характеристик от штатного режима при доле ложных срабатываний меньшей, чем можно получить простым уменьшением порога ε_F в выражении $|F(t) - F_0(t)| > \varepsilon_F$.

В третьей главе построена объектная модель системы мониторинга. На основе предложенной модели определены принципиальные свойства взаимодействующих объектов, предложен шаблон описания сложной системы, позволяющий формализовать характерные для системы задачи и применить к ним предложенные методы и алгоритмы.

- Разработана объектная модель системы мониторинга и приведены примеры её применения для описания систем мониторинга в области эпидемиологии и транспортной безопасности;
- Рассмотрены классический и альтернативный предметно-независимый подходы к построению алгоритмов мониторинга;
- Произведён обоснованный выбор формальной модели представления знаний предметной области для использования в предметно-независимых алгоритмах;
- Разработана схема системы мониторинга, использующей предметно-независимые алгоритмы интеллектуальной обработки данных, рассмотрены:
 - Вопросы обработки входных данных системы мониторинга;
 - Способ оценки неопределённости характеристик объектов мониторинга, используемый в интеллектуальных алгоритмах обработки данных;
 - Алгоритм идентификации и устранения противоречивых информационных сообщений;
 - Алгоритм кластеризации объектов мониторинга для формирования гипотезы о характере внешнего воздействия;

В разделе 3.1 разработана объектная модель системы мониторинга, применяемая при разработке информационных систем мониторинга. В формулировках задач мониторинга, приведённых в разделе 2.4, определены абстрактные элементы модели, которые необходимо связать с концепциями предметной области при проектировании той или иной системы мониторинга: множество объектов W , вектор состояния объекта \mathbf{D} , классы объектов C_1, \dots, C_n , функцию $F(t)$ агрегированных характеристик системы W , функцию $F_0(t)$ штатного режима функционирования системы W , норму $|\cdot|$ и порог ε_F отклонения $F(t)$ от штатной функции $F_0(t)$, а также альтернативные гипотезы $H = \{H_1, \dots, H_N\}$ и пороговое значение ε_H .

Диагностирование состояния $\mathbf{D}(t)$ объекта w системы производится при помощи источников информации S , результатом работы которых $\widehat{\mathbf{D}}(t)$ является сообщение об изменении состояния объекта:

$$E(t_j) = \begin{cases} 1: \rho(\widehat{\mathbf{D}}(t_j) - \widehat{\mathbf{D}}(t_{j-1})) > \varepsilon_D, \\ 0: \text{иначе} \end{cases}$$

где функция $E(t_j)$ принимает значение 1 в те моменты времени t_j , когда обнаружено значительное изменение свойств объекта; t_j – моменты времени, в которые производились измерения состояния объекта; ε_D – порог, характеризующий значимость изменения. Информацию о событии E несёт сообщение $\widehat{\mathbf{D}}(t)$, производимое источником информации, поэтому наблюдаемое время t_j возникновения события зависит от принципа работы источников информации и их набора. Изменение состояния $\widehat{\mathbf{D}}$ объекта w может быть зафиксировано несколькими источниками информации в разные моменты времени. Каждый источник информации передаст сообщение о событии, в результате чего получим последовательность сообщений об одном и том же событии: $E(t_j) = 1, E(t_{j+1}) = 1, \dots, E(t_{j+n}) = 1$, где n – количество источников информации, обнаруживших изменение состояния $\widehat{\mathbf{D}}(t)$ объекта. Существует несколько причин, которые определяют «размножение» события, среди них:

- Состояние $\widehat{\mathbf{D}}$ объекта w фиксируют несколько источников информации: S_1, S_2, S_3, \dots ;
- Источники S_i и S_j фиксируют разные характеристики объекта w , то есть получают разный набор входных данных: $\mathbf{D} = \{\mathbf{D}_i, \mathbf{D}_{ij}, \mathbf{D}_j\}$, где \mathbf{D}_i – информация источника S_i , \mathbf{D}_j – информация источника S_j , \mathbf{D}_{ij} – информация, измеренная обоими источниками;
- Источники S_i и S_j производят разный набор результирующих значений: $\widehat{\mathbf{D}} = \{\widehat{\mathbf{D}}_i, \widehat{\mathbf{D}}_{ij}, \widehat{\mathbf{D}}_j\}$, где $\widehat{\mathbf{D}}_i$ – множество значений источника S_i , $\widehat{\mathbf{D}}_j$ – множество значений источника S_j , $\widehat{\mathbf{D}}_{ij}$ – множество значений, общее для источников S_i и S_j ;
- События E_i и E_j похожим образом изменяют вектор состояния объекта $\widehat{\mathbf{D}}$.

В результате, вместо одного момента времени t для каждого события получаем n значений, что искажает наблюдаемую функцию $F(t)$ и, как следствие, приводит к сбоям при обнаружении отклонения состояния системы от штатного режима и при прогнозировании поведения системы. Решение проблемы искажения функции $F(t)$, возникающего вследствие «размножения» сообщений о событии, связано с решением двух задач:

1. Идентификации повторных и противоречивых сообщений: Для сообщений о событиях $E(t_i)$ и $E(t_j)$ установить, сигнализируют они об одном и том же событии, или о разных;
2. Кластеризации сообщений по их принадлежности к различным протекающим в системе информационным процессам.

Проблема характерна для систем мониторинга, в которых состояние многокомпонентной системы определяется множеством взаимосвязанных источников информации. На рис. 2 представлена UML-диаграмма классов объектной модели такой системы. Объектная модель используется при проектировании программного обеспечения прикладных систем мониторинга. Соответствие между UML-диаграммой и математической моделью задаётся в табл. 1.

Приведены примеры соответствия объектов модели и предметной области, а также сформулированы задачи идентификации повторных и противоречивых сообщений и кластеризации сообщений по их принадлежности к информационным процессам в терминах предметной области.

Раздел 3.2. Для обнаружения и исправления систематических ошибок измерения состояния $F(t)$ системы W требуется привлечение знаний предметной области. Это может быть достигнуто как разработкой специальных алгоритмов для узких предметных областей, так и использованием обучаемых интеллектуальных структур. В первом случае тиражируемость решения ограничена, так как алгоритмы включают специфику предметной области и наработки передаются только с опытом разработчика. При использовании обучаемых интеллектуальных структур требуется значительное количество обучающих примеров. Их разработка должна выполняться



Рис. 2. UML-диаграмма классов объектной модели системы мониторинга

Таблица 1. Соответствие элементов UML-диаграммы и переменных математической модели.

1. Наблюдаемая система	W
2. Объект	w
3. Процесс	$F(t)$
4. Событие	E
5. Источник информации	S
5.1. Измеряемые параметры	D
5.2. Множество значений измеряемых параметров	
6. Сообщение	$E(t)$
6.1. Множество значений выходных параметров	\hat{D}

совместно специалистами предметной области и специалистами в области искусственного интеллекта. В результате обученная модель также применяется только к решению узкой задачи.

Проблемой описанных методов является то, что формализованные знания предметной области заложены в алгоритмы обработки информации, что делает невозможным их простую адаптацию к новым задачам. В работе предложены алгоритмы обработки данных, которые взаимодействует с формализованной моделью знаний предметной области, для решения задач, определяемых данной моделью (рис. 3).

На рисунке введены следующие обозначения: А – алгоритм обработки данных; D – входные данные алгоритма; D^{\sim} - выходные данные алгоритма; DM – интерфейс входных данных (модель данных); К – знания предметной области; КМ – интерфейс модели знаний. Принципиальным отличием предлагаемой структуры алгоритмов интеллектуального анализа и обработки информации является выделение модели знаний предметной области в независимый блок и взаимодействие алгоритма с этим блоком средствами формализованного интерфейса модели знаний.

В разделе 3.3 произведён обоснованный выбор формальной модели представления знаний предметной области. Модель должна отражать существенные свойства моделируемого объекта или явления, поэтому требования к модели

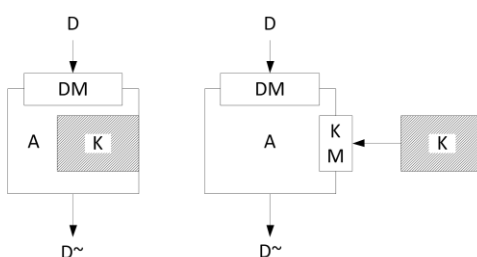


Рис. 3. Традиционные алгоритмы (слева) и новые алгоритмы с двумя входами (справа)

представления знаний также определяются свойствами моделируемого объекта – знаний. Когда говорят об отличиях знаний от информации, указывают следующие свойства: интерпретируемость, активность, связность, структурированность. Типовые модели представления знаний в той или иной степени отражают перечисленные свойства знаний, но ни одна из типовых моделей знаний не является универсальным решением. Однако различные модели знаний имеют общие черты, что позволяет строить гибридные модели знаний на основе типовых. Поэтому, задача выбора формальной модели представления знаний является задачей синтеза модели, удовлетворяющей требованиям поставленной задачи на основе типовых.

Предложена система критериев для выбора формальной модели представления знаний. На верхнем уровне критерии поделены на три группы: «Свойства представления», «Эксплуатационные характеристики» и «Удобство использования».

В группу «Свойства представления» входят критерии, описывающие принципиальные свойства модели представления знаний: способность к обучению, статические характеристики, характеристики методов вывода. К эксплуатационным характеристикам относятся: временная сложность, ресурсоёмкость, возможность распараллеливания. Группа характеристик верхнего уровня «Удобство использования» включает такие критерии, как сложность разработки, сложность восприятия.

В разрезе предложенной системы характеристик рассмотрены типовые модели представления знаний, выбрана базовая модель и предложены доработки, позволяющие использовать её в предметно-независимых алгоритмах мониторинга. В качестве базовой выбрана модель семантической сети, для работы с неточными знаниями предложено использовать методы нечёткой логики, а вероятностные знания и обучаемость структуры обеспечить засчёт аппарата Байесовских сетей.

В разделе 3.4 разработана схема системы мониторинга, использующей предметно-независимые алгоритмы интеллектуальной обработки данных. На рис. 4 представлена общая схема системы мониторинга, пунктирной линией выделена часть, в которой предметно-независимым способом решаются задачи идентификации повторных и противоречивых сообщений и кластеризации сообщений по их

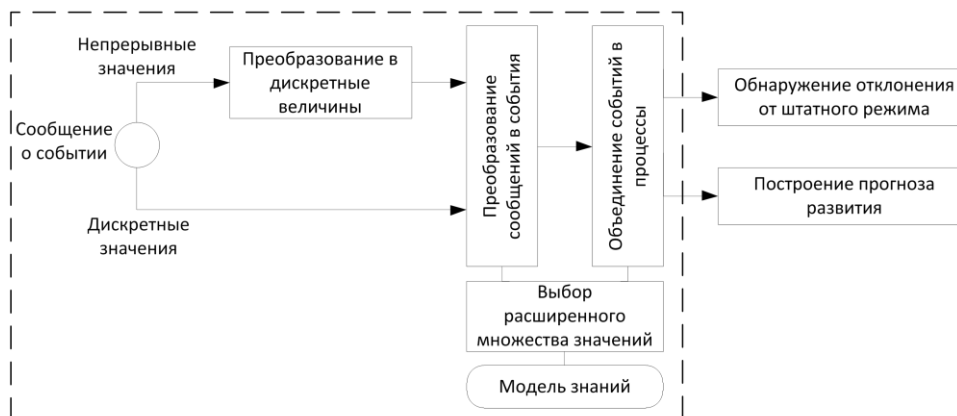


Рис. 4. Общая схема решения

принадлежности к информационным процессам.

Приведено описание процедур предметно-независимого блока. Описаны процедуры, обозначенные на схеме, а также алгоритмы и модели, необходимые для этих процедур:

1. Преобразование непрерывных входных параметров в дискретные значения;
2. Мера близости векторов состояний объектов;
3. Выбор формальной модели представления знаний предметной области;
4. Алгоритм выбора расширенного множества значений состояния объекта;
5. Преобразование сообщений в события;
6. Объединение событий в процессы;

В разделе 3.4.1 рассмотрены вопросы преобразования непрерывных входных параметров в дискретные значения. Данные о состоянии объекта \mathbf{D} должны быть представлены в формализованном виде. Для этого должны быть выбраны шкалы и единицы измерения. В случае, когда элемент d_i вектора \mathbf{D} является численным результатом измерения, то единицы измерения и шкалы известны. Если для описания значений элемента d_i используются лингвистические конструкции, то требуется их формализация.

Наблюдаемые параметры могут быть описаны непрерывными или дискретными значениями. И вектор значений \mathbf{D} состоит из дискретной и непрерывной составляющих: $\mathbf{D} = \{\mathbf{D}^{\text{дискр.}}, \mathbf{D}^{\text{непр.}}\}$.

Дискретным значениям ставятся в соответствие узлы семантической сети. Для этого задаётся отображение множества значений каждой из составляющих $d_i \in \mathbf{D}^{\text{дискр.}}$ дискретной составляющей вектора \mathbf{D} на узлы семантической сети $V: D_i \rightarrow V$, где D_i – счётное и конечное множество значений i -ой компоненты d_i вектора $\mathbf{D}^{\text{дискр.}}$. В общем случае отображение может не быть взаимно однозначным. Например, нескольким элементам множества D_i может соответствовать один элемент множества V , что возможно, когда используется недостаточно подробная семантическая сеть, либо значения множества D_i содержит «синонимы»: $\exists d_j, d_k \in D_i: d_j \approx d_k$.

Отображение непрерывных значений $\mathbf{D}^{\text{непр.}}$ на узлы семантической сети требует дополнительного шага. Непрерывные значения должны быть поделены на диапазоны, каждый из которых соответствует узлу семантической сети. Для представления диапазонов используются нечёткие множества, характеристическая функция $\mu(d_i^{\text{непр.}})$ которых зависит от решаемой задачи.

В разделе 3.4.2 вводится мера близости векторов состояний объектов. Сообщение о событии содержит информацию $\hat{\mathbf{D}}(t)$ о состоянии объекта в момент времени t , поэтому, мера близости сообщений зависит от меры близости векторов состояний объектов. Введём меру близости $\rho(\mathbf{D}_i, \mathbf{D}_j)$ векторов состояний объектов $\mathbf{D}_i(t)$ и $\mathbf{D}_j(t)$ и выберем алгоритм её измерения.

Состояние объекта $\hat{\mathbf{D}}$ описывается наблюдаемой $\hat{\mathbf{D}}^{\text{набл.}}$ и скрытой $\hat{\mathbf{D}}^{\text{скр.}}$ частями, между которыми существует вероятностная связь:

$$\hat{\mathbf{D}} = \{\hat{\mathbf{D}}^{\text{набл.}}; \hat{\mathbf{D}}^{\text{скр.}}\}, P(\hat{\mathbf{D}}^{\text{скр.}}) = P(\hat{\mathbf{D}}^{\text{скр.}} | \hat{\mathbf{D}}^{\text{набл.}}),$$

где $P(\hat{\mathbf{D}}^{\text{скр.}})$ – вероятность того, что скрытое состояние $\hat{\mathbf{D}}^{\text{скр.}}$ соответствует действительному состоянию объекта $\mathbf{D}^{\text{скр.}}$. При этом считаем, что наблюдаемое состояние объекта $\mathbf{D}^{\text{набл.}}$ измерено без ошибок, т.е. $P(\hat{\mathbf{D}}) = P(\hat{\mathbf{D}}^{\text{скр.}} | \hat{\mathbf{D}}^{\text{набл.}}) \cdot P(\hat{\mathbf{D}}^{\text{набл.}}) = P(\hat{\mathbf{D}}^{\text{скр.}} | \hat{\mathbf{D}}^{\text{набл.}})$.

Таким образом, задача сводится к измерению близости векторов состояний $\hat{\mathbf{D}}^{\text{скр.}}$ при условии получения результатов измерения $\hat{\mathbf{D}}^{\text{набл.}}$. Введём меру близости векторов $\hat{\mathbf{D}}_i^{\text{скр.}}$ и $\hat{\mathbf{D}}_j^{\text{скр.}}$ в следующем виде:

$$\rho(\hat{\mathbf{D}}_i^{\text{скр.}}, \hat{\mathbf{D}}_j^{\text{скр.}}) = \min[P(\hat{\mathbf{D}}_i^{\text{скр.}} | \hat{\mathbf{D}}^{\text{набл.}}); P(\hat{\mathbf{D}}_j^{\text{скр.}} | \hat{\mathbf{D}}^{\text{набл.}})],$$

следовательно,

$$\rho(\mathbf{D}_i, \mathbf{D}_j) = \rho(\mathbf{D}_i^{\text{скр.}}, \mathbf{D}_j^{\text{скр.}}) = \min[P(\mathbf{D}_i^{\text{скр.}} | \mathbf{D}^{\text{набл.}}); P(\mathbf{D}_j^{\text{скр.}} | \mathbf{D}^{\text{набл.}})].$$

На практике данная мера близости будет мала для тех пар векторов состояний $\bar{\mathbf{S}}$, скрытые части $\bar{\mathbf{S}}^{\text{скр.}}$ которых с близкой вероятностью проявляются при наблюдаемых измеренных значениях $\bar{\mathbf{S}}^{\text{набл.}}$.

Предложен алгоритм выбора расширенного множества значений состояния объекта. Использование графа семантической сети в качестве основы построения Байесовской сети позволяет свести процедуру обучения к выбору значений условных вероятностей для рёбер графа сети, имея в наличии набор узлов и связей между ними. Таким образом, задача обучения семантической сети разделяется на две подзадачи:

1. Первоначальное обучение, включающее выбор онтологии предметной области и её подстройку при помощи экспертов;
2. Корректировка вероятностных коэффициентов Байесовской сети на основе данных, появляющихся при использовании созданной с использованием системы мониторинга.

Первоначальное обучение семантической сети начинается с выбора графа G , представляющего онтологию предметной области. Необходимым условием применения графа G , представляющего онтологию предметной области, для оценки меры близости векторов является возможность сопоставления $\mathbf{D}^{\text{скр.}} \rightarrow C^{\text{скр.}}$ и $\mathbf{D}^{\text{набл.}} \rightarrow C^{\text{набл.}}$. То есть вершины графа G должны содержать множество, соответствующее наблюдаемым и скрытым параметрам и их дискретным значениям.

Граф семантической сети должен выражать причинно-следственные связи между концептами, поэтому необходимым шагом подготовки графа онтологии является устранение циклов. Результатом устранения циклов станет граф, являющийся суперпозицией древовидных графов: $G = T_1 \cup T_2 \cup \dots$

При использовании в качестве основы для построения графа G онтологии предметной области, отдельные деревья T_i в составе графа G представляют ни что иное, как таксономии, элементами которых являются элементы множеств $C^{\text{скр.}}$, $C^{\text{набл.}}$, а также элементы, соответствующие промежуточным уровням классификации.

Назначение условных вероятностей рёбрам графа происходит следующим образом. Предположим, что $c \in C^{\text{скр.}}$, а $c_i \in C^{\text{набл.}}$, $i = 1 \dots n$ – узлы, соответствующие «причинам» узла c . В случае если наблюдаемые параметры c_i попарно независимы, то вероятность события c определяется по формуле полной вероятности:

$$p(c) = \sum_i \mathbf{D}^{\text{набл.}}[i] \cdot p(c|c_i),$$

где $\mathbf{D}^{\text{набл.}}[i]$ равняется 1, если i -ый параметр наблюдается и 0 иначе. В случае, когда между наблюдаемыми параметрами c_i существует зависимость, обучение потребует записи условной вероятности $p(c|\cup c_i)$, где $\cup c_i$ – всевозможные комбинации событий c_i .

При использовании данного подхода требуется значительное количество примеров для обучения сети, которое определяется числом перестановок элементов c_i для каждого элемента $c \in C^{\text{скр.}}$. Сложность возрастает, если принять во внимание тот факт, что состояние некоторых параметров $c_i \in C^{\text{набл.}}$ может приобретать не только два значения – 1 или 0, но и значение «неизвестно».

С другой стороны, граф G представлен в виде суперпозиции древовидных графов T_i , при этом T_i являются таксономиями, то есть классификациями концептов сходными признаками. Это обозначает, что чем меньше расстояние между узлами дерева T , тем ближе расположены соответствующие векторы состояний \mathbf{D} : $\rho(\mathbf{D}_i^{\text{скр.}}, \mathbf{D}_j^{\text{скр.}}) \sim l(C^{\text{скр.}}_i, C^{\text{скр.}}_j)$, где $l(C^{\text{скр.}}_i, C^{\text{скр.}}_j)$ – расстояние между узлами графа $C^{\text{скр.}}_i, C^{\text{скр.}}_j$, соответствующим скрытым состояниям $\mathbf{D}_i^{\text{скр.}}, \mathbf{D}_j^{\text{скр.}}$.

В дальнейшем, мера близости $\rho(\mathbf{D}_i^{\text{скр.}}, \mathbf{D}_j^{\text{скр.}})$ будет использоваться для кластеризации состояний $\mathbf{D}^{\text{скр.}}$, то есть необходимо выбрать параметр ε такой, что при $\rho(\mathbf{D}_i^{\text{скр.}}, \mathbf{D}_j^{\text{скр.}}) < \varepsilon$, векторы $\mathbf{D}_i^{\text{скр.}}, \mathbf{D}_j^{\text{скр.}}$ попадают в один кластер. При этом деревья T в составе графа G , основанного на онтологии, представляют таксономию – классификацию концептов. То есть, чем ближе находится общий родитель двух концептов, тем более близки векторы $\mathbf{D}^{\text{скр.}}$, которые они представляют.

Таким образом, для первоначального обучения семантической сети при помощи эксперта меру близости векторов $\rho(\mathbf{D}_i^{\text{скр.}}, \mathbf{D}_j^{\text{скр.}})$ можно записать в виде:

$$\rho(\mathbf{D}_i^{\text{скр.}}, \mathbf{D}_j^{\text{скр.}}) = l(C^{\text{скр.}}_i, C^{\text{скр.}}_j) = \min(l(C^{\text{скр.}}_i, C_p), l(C^{\text{скр.}}_j, C_p)),$$

где C_p – общий «родитель» узлов $C^{\text{скр.}}_i$ и $C^{\text{скр.}}_j$. В случае, когда общего родителя нет, $\rho(\mathbf{D}_i^{\text{скр.}}, \mathbf{D}_j^{\text{скр.}}) = \infty$. При определении общего родителя $C^{\text{скр.}}_i$ и $C^{\text{скр.}}_j$ поиск производится только на множестве $E(C^{\text{скр.}}_i, C^{\text{скр.}}_j) = E(C^{\text{скр.}}_i) \cap E(C^{\text{скр.}}_j)$, где $E(C^{\text{скр.}}_i)$ – расширенное множество значений элемента $C^{\text{скр.}}_i$, определяемое как множество

наследников узла, находящегося на n узлов выше $C^{\text{скр.}}$: $E(C^{\text{скр.}}) = \text{succ}(\text{par}(C^{\text{скр.}}, n))$, где $\text{par}(C, n)$ – множество всех родителей узла C , находящихся n узлами выше. Множество может иметь размерность больше 1, так как узел $C \in G$ может входить в несколько деревьев T ; $\text{succ}(C)$ – множество всех «наследников» узла C во всех деревьях T в которые он входит.

Число n для каждого узла C определяется для каждого вхождения узла C в дерево T на основании экспертной оценки, либо в зависимости от связей узла $C^{\text{скр.}}$ с узлами $C^{\text{набл.}}$: $P(\bar{C}^{\text{скр.}} | \bar{C}^{\text{набл.}})$.

Экспертная оценка заключается в выборе эмпирической зависимости числа n от значений вектора состояния $D^{\text{набл.}}$: $n = n(D^{\text{набл.}})$. Использование экспертной оценки числа n позволяет использовать онтологию предметной области в качестве основы для Байесовской сети в случае, когда для оценки условных вероятностей $P(D^{\text{скр.}} | D^{\text{набл.}})$ не существует достаточно данных для обучения сети.

В разделе 3.4.3 разработан алгоритм идентификации и устранения противоречивых информационных сообщений. Расстояние $\rho(D_i^{\text{скр.}}, D_j^{\text{скр.}})$ между векторами состояний $D^{\text{скр.}}$, введённое указанным выше способом, позволяет установить меру близости состояний D_i, D_j объектов w . Сообщение о событии несёт также информацию о времени события и данные I , идентифицирующие источник события - объект w : $\{I, D, t\}$. В случае, когда сообщения D_i, D_j несут информацию о состоянии одного и того же объекта w , то есть $\rho(I_i, I_j) < \theta$, возникает задача идентификации и устранения повторяющихся и противоречивых сообщений о состоянии объекта многокомпонентной системы. В противном случае, решается задача кластеризации данных для идентификации независимых процессов, протекающих в системе.

При поступлении сообщения, вектору скрытого состояния $D^{\text{скр.}}$ ставится в соответствие расширенное множество значений E , величина которого в соответствии с алгоритмом выбора расширенного множества значений тем больше, чем меньше вероятность $P(D^{\text{скр.}} | D^{\text{набл.}})$:

$$|E| \sim n \sim \frac{1}{P(D^{\text{скр.}} | D^{\text{набл.}})}$$

Величина расширенного множества значений $|E|$ характеризуют дисперсию измеренного состояния объекта $D^{\text{скр.}}$. Сообщения, для которых $|E|$ превышает заданное пороговое значение, считаются недостоверными.

При поступлении двух или более сообщений D_i, D_j , расширенные множества значений которых пересекаются, принимается решение об идентичности сообщений. Устраняются все идентичные сообщения, кроме одного. Его выбор определяется либо минимальной дисперсией, либо свойствами, зависящими от условий задачи –

например, самое новое. Сообщение, представляющее множество идентичных сообщений, а также его дисперсию, будем называть событием.

В разделе 3.4.4 разработан алгоритм кластеризации объектов мониторинга для формирования гипотезы о характере внешнего воздействия. События представлены сообщениями, для которых $\rho(I_i, I_j) > \theta$, то есть сообщения несут информацию о разных объектах w_i и w_j . Процессом считается множество событий, связанных общими свойствами. Для объединения сообщений в процессы также используется алгоритм выбора расширенного множества значений с незначительными дополнениями.

Решение об объединении событий в процесс принимается при выполнении следующих условий:

$$\begin{cases} \rho(\mathbf{D}_i, \mathbf{D}_j) > \theta \\ E(\mathbf{D}^{\text{скр.}}_i, \mathbf{D}^{\text{скр.}}_j) \neq \emptyset. \\ \rho(t_i, t_j) < \tau \end{cases}$$

Последнее условие отражает близость событий по времени, при этом τ – характерное время события, зависящее от вида события $\tau = \tau(\mathbf{D}^{\text{скр.}})$.

При объединении событий в процессы можно использовать меру близости $\rho_{\{T\}}(\mathbf{D}_i^{\text{скр.}}, \mathbf{D}_j^{\text{скр.}})$, которая отличается от описанной выше меры $\rho(\mathbf{D}_i^{\text{скр.}}, \mathbf{D}_j^{\text{скр.}})$ тем, что вычисление происходит только по подмножеству $\{T\}$ деревьев, составляющих граф G .

В четвёртой главе приводятся результаты применения предложенных моделей и методов при разработке конкретных информационных систем – описано приложение предложенных процедур и алгоритмов к проектированию и разработке систем мониторинга распространения вирусного гепатита и планирования контрольных мероприятий в сфере транспортной безопасности. Приводятся:

- Цели и задачи создания информационных систем мониторинга распространения вирусного гепатита и планирования контрольных мероприятий в сфере транспортной безопасности;
- Описание предметной области;
- Результаты анализа требований к разрабатываемым информационным системам мониторинга;
- Проекты рассматриваемых информационных систем мониторинга;

Система мониторинга распространения вирусного гепатита предназначена для повышения эффективности эпидемиологического надзора за заболеваемостью вирусными гепатитами: оперативного контроля заболеваемости, систематизации процессов сбора, хранения и обработки больших объёмов первичных данных, связанных с заболеваемостью, анализа накопленных данных по заболеваемости с помощью методов и средств математической статистики. Анализ структуры

заболеваемости, позволяет органам Роспотребнадзора своевременно и обоснованно принимать управленческие решения по обеспечении противоэпидемиологических и профилактических мероприятий.

Процесс эпидемиологического мониторинга требует совместной работы множества специалистов, отличающихся выполняемыми функциями и полномочиями. В процессе анализа процессов предметной области были выявлены основные участники процесса эпидемиологического мониторинга: медработники лечебно-профилактических учреждений, врачи-лаборанты лабораторий, врачи-эпидемиологи четырёх уровней иерархии.

Диаграмма вариантов использования (рис. 5) демонстрирует необходимость совместной работы различных участников процесса при выполнении типовых задач, возникающих в процессе эпидемиологического мониторинга.

Взаимодействие множества специалистов, вовлечённых в процесс эпидемиологического мониторинга, приводит к появлению следующих нежелательных свойства данных мониторинга:

- Противоречивость и наличие повторов, возникающие вследствие частичного дублирования функций специалистов;
- Неоднородность представления, возникающая вследствие применения специалистами различных методик наблюдения одного и того же явления.

Предложенные в работе методы позволяют устранять эти свойства, помогая участникам процесса эпидемиологического мониторинга исправлять некорректные данные и обнаруживать скрытые взаимосвязи между данными мониторинга, полученными разными специалистами. На рис.6 изображён элемент экрана системы мониторинга распространения вирусных гепатитов, демонстрирующий обнаружение повторов в первичных данных мониторинга.

Пользователь имеет возможность согласиться или отказаться от группировки



Рис. 5 Диаграмма вариантов использования для типовых задач, выполняемых в процессе эпидемиологического мониторинга.

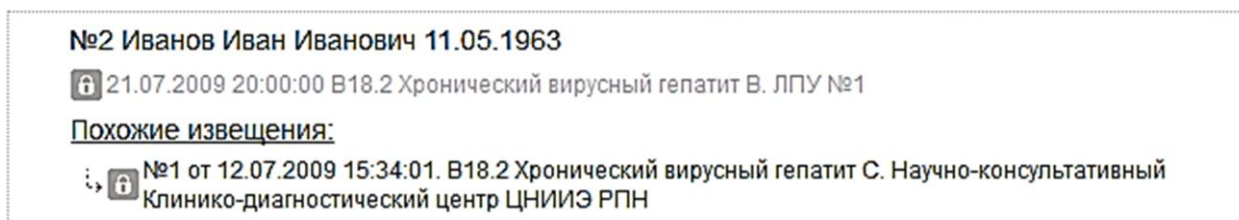


Рис. 6 Элемент экрана системы мониторинга распространения вирусных гепатитов, демонстрирующий обнаружение повторов в первичных данных мониторинга

экстренных извещений, предложенной системой. Отказ пользователя от группировки позволяет скорректировать параметры алгоритма группировки экстренных извещений.

В качестве основы информационной модели предметной области, использованной при разработке системы мониторинга распространения вирусных гепатитов, была выбрана семантическая сеть, входящая в состав международного медицинского классификатора UMLS (Unified Medical Language System).

Предложенные в работе методы были также применены при разработке системы планирования контрольных мероприятий в сфере транспортной безопасности. Система предназначена для консолидации информации о результатах контрольной деятельности Ространснадзора, анализа результатов контрольных мероприятий и составления плана будущих контрольных мероприятий.

Формально, процесс составления плана контрольных мероприятий является задачей многокритериальной оптимизации, однако необходимость учёта особенностей конкретных транспортных предприятий затрудняет формализацию критериев планирования. Поэтому, процесс планирования представляет собой последовательность согласований, выполняемых территориально распределёнными подразделениями Ространснадзора.

Целью процесса планирования является составление допустимого варианта плана мероприятий. Допустимость плана определяется разрешённой частотой проверок по заданной тематике, возможностью проведения одновременных проверок одной организации, наличием ресурсов для проведения проверки, тематикой и результатами предыдущих проверок и т.д.

Применение предложенных в работе методов мониторинга позволяет пользователям системы оперативно обнаруживать конфликтующие контрольные мероприятия, нарушающие условия допустимости плана, что повышает оперативность процесса планирования. Использование предложенной в работе информационной модели предметной области позволяет подключать ряд новых критериев планирования без изменения программного кода системы, что снижает материальные и временные затраты, необходимые для адаптации системы планирования контрольных мероприятий к изменяющимся требованиям пользователей.

В приложениях представлены документы, подтверждающие внедрение результатов диссертационной работы, описание структур разработанных информационных систем мониторинга.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

1. Разработаны формальная теоретико-множественная модель многокомпонентной системы мониторинга, а также объектная модель для проектирования информационных систем, в которой специфика предметной области изолирована в формальной модели знаний предметной области;
2. Проведён анализ существующих формальных моделей представления знаний предметной области. На основании проведённого анализа выбрана формальная модель представления знаний предметной области для использования в интеллектуальных алгоритмах систем мониторинга;
3. Разработаны методы интеллектуального анализа и обработки экспертной информации с использованием формальной модели знаний предметной области, включая способ оценки неопределённости характеристик объектов мониторинга и алгоритм кластеризации объектов мониторинга для формирования гипотезы о характере внешнего воздействия;
4. Разработаны проект и прототип программного комплекса, включающего в себя элементы перечисленных моделей и алгоритмов, осуществляющего сбор экспертных оценок состояния объектов и их интеллектуальную обработку с использованием проблемно-ориентированных семантических графов для систем мониторинга распространения вирусного гепатита и планирования контрольных мероприятий в сфере транспортной безопасности.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИОННОЙ РАБОТЫ

1. Баканова Н.Б., Гурвиц А.Л., Цапаева Ю.А., Сурпин В.П. Создание корпоративной системы сбора и анализа информации на основе Веб-технологий // Труды Международного семинара «Распределенные компьютерные и телекоммуникационные сети». М.: ИППИ РАН, 2005. С. 193-196.
2. Сурпин В.П., Волчков Д. В. Проектирование структур данных для комплекса сбора оперативной информации о случаях заболеваний // Информационные технологии и системы: Сборник трудов. М.: ИППИ РАН, 2008. С. 126-128.
3. Сурпин В.П. Проектирование комплекса сбора оперативной информации о случаях заболеваний // III Всероссийская молодёжная конференция по проблемам управления: Сборник трудов. М.: ИПУ РАН, 2008. С. 283-284.
4. Сурпин В.П. Разработка подсистемы ведения классификаторов корпоративной информационной системы // «Информационные технологии». 2009. №11. С. 23-27.
5. Сурпин В.П. Алгоритм анализа входных данных для крупномасштабной системы сбора медицинской статистики // Управление развитием крупномасштабных систем: Материалы Четвёртой международной конференции. М.: ИПУ РАН, 2009. Т. 2. С. 283-284.

6. Сурпин В.П.. An approach to data validation based on lifecycle-bounded metadata // Proceedings of the 4th Spring Young Researchers' Colloquium on Software Engineering. М.: ИСП РАН, 2009. С. 33-36.
7. Сурпин В.П. Нефункциональные требования и методология быстрой разработки приложений. Подход к прогнозированию производительности // Управление развитием крупномасштабных систем: Материалы Четвёртой международной конференции. М.: ИПУ РАН, 2010. Т. 2. С. 204-206.
8. Сурпин В.П. Применение методов сетей массового обслуживания в задаче распределения ресурсов «облачных» систем // Информационные технологии и системы – 2011: Сборник трудов. М.: ИППИ РАН, 2011. С. 227-232
9. Сурпин В.П. Метод тестирования гипотез о динамике информационных процессов в сложных системах // Обозрение прикладной и промышленной математики. 2011. Т. 18, Вып. 2. С. 334.
10. Сурпин В.П. Мониторинг многокомпонентных систем: предметно-независимые модели и методы // Информационные процессы. 2011. Т. 11. Вып. 3. С. 378-393.

