

УДК 519.72

АНАЛИЗ ИНФОРМАЦИИ И ПРИНЯТИЕ РЕШЕНИЙ В СИСТЕМАХ ИНФОРМАЦИОННОГО МОНИТОРИНГА

Лебедев Анатолий Анатольевич¹, Рыжов Александр Павлович²

¹ Аспирант;

МГУ им. М.В. Ломоносова, механико-математический факультет;
119234, г. Москва, Воробьевы Горы, ГЗ МГУ;
e-mail: anatoli.lebedev@gmail.com.

² Доктор технических наук, доцент;

МГУ им. М.В. Ломоносова, механико-математический факультет;
119234, г. Москва, Воробьевы Горы, ГЗ МГУ;
e-mail: alexander.ryjov@gmail.com.

Технология информационного мониторинга была разработана для анализа сложных, слабоформализованных проблем (процессов) на основе всей доступной информации, построения прогнозов их развития и выработки рекомендаций по управлению их развитием. В работе описываются основные практические аспекты разработки систем информационного мониторинга, приводятся теоретические результаты, гарантирующие оптимальность их работы, и описываются примеры применения технологии в различных предметных областях.

Ключевые слова: информационный мониторинг, нечеткие иерархические системы, принятие решений.

INFORMATION ANALYSIS AND DECISION MAKING IN INFORMATION MONITORING SYSTEMS

Lebedev Anatoly¹, Ryjov Alexander²

¹ Postgraduate student;

Lomonosov Moscow State University, faculty of Mechanics & Mathematics;
119234, Moscow, Vorobyevy Gory, Glavnoe Zdanie MGU;
e-mail: anatoli.lebedev@gmail.com.

² Doctor of Science in Engineering, associate professor;

Lomonosov Moscow State University, faculty of Mechanics & Mathematics;
119234, Moscow, Vorobyevy Gory, Glavnoe Zdanie MGU;
e-mail: alexander.ryjov@gmail.com.

Information monitoring technology has been developed for analyzing complex systems that do not have and appropriate mathematical model. In this paper, we discuss the principal aspects of developing such systems, provide some theoretical results that guarantee their optimality, and describe several examples of applying the technology for various application domains.

Keywords: information monitoring, fuzzy hierarchical systems, decision making.

Введение

Содержание работы многих классов специалистов в современном информационном обществе можно представить как работу по определенными проблемам. В наиболее общем виде она заключается в оценке текущего состояния проблемы (процесса) на основе всей доступной информации, прогнозировании дальнейшего хода событий и выработке рекомендаций по управлению развитием про-

блемы исходя из целей, стоящих перед специалистом. Можно привести следующие примеры проблем из различных областей:

- моделирование поведения клиента (маркетинг);
- сегментирование рынков (маркетинг);
- диагностика (медицина);
- повышение эффективности работы предприятия (менеджмент).

Свойство перечисленных проблем – «слабая формализация» (наличие как количественных, так и качественных признаков, отсутствие математических моделей), структурная организация (проблема/процесс имеет некоторую структуру) и наличие человека как активного элемента системы. Эти свойства препятствуют автоматизации процесса решения этих задач классическими средствами. С другой стороны, цена ошибки во многих таких задачах очень высока, поэтому очень желательно снизить влияние человеческого фактора.

Именно с этими трудностями призвана бороться технология информационного мониторинга, разработанная для анализа сложных, слабоформализованных проблем (процессов) на основе всей доступной информации, построения прогнозов их развития и выработки рекомендаций по управлению их развитием. Системы информационного мониторинга, разработанные на базе этой технологии, ориентированы на обработку информации, обладающей такими свойствами как:

- разнородность, т.е. фиксация информации в виде статей, газетных заметок, компьютерном виде, аудио- и видеоинформация и т.п.;
- фрагментарность: информация чаще всего относится к какому-либо фрагменту проблемы, причем разные фрагменты могут быть по-разному «покрыты» информацией;
- разнородность: информация может относиться ко всей проблеме в целом, к некоторой ее части, к конкретному элементу проблемы;
- различная степень надежности.

Системы, разработанные на базе этой технологии, позволяют иметь развивающуюся во времени модель проблемы на основе оценок аналитиков, подкрепленную ссылками на все информационные материалы, выбранные ими, с общими и частными оценками состояния проблемы и/или ее аспектов. Использование времени как параметра системы позволяет проводить как ретроспективный анализ, так и строить прогнозы развития проблемы. В последнем случае также возникает возможность выделения «критических путей» – таких элементов (или наборов элементов) модели, небольшое изменение которых может вызвать заметные изменения в состоянии всей проблемы. Знание таких элементов позволяет выявить «слабые места» в проблеме на текущий момент времени, разработать мероприятия по блокированию нежелательных ситуаций или провоцированию желательных, т.е. в некоторой степени управлять развитием проблемы в интересах организации, ее отслеживающей.

1. Основные этапы разработки и использования систем информационного мониторинга

Можно выделить следующие стадии разработки систем информационного мониторинга:

1. *Построение структурной модели* предметной области, то есть выделение факторов, играющих роль в исследуемой области, и представление их в виде иерархической структуры.
2. *Построение поведенческой модели* предметной области, то есть задание зависимостей между факторами, выделенными на первом этапе.
3. *Инициализация* – сбор информации о состоянии параметров системы и введение данных в модель.

Ниже эти стадии будут описаны подробнее.

1.1. Создание структурной модели

Структурная модель представляет собой иерархически организованный набор факторов, определяющих состояние наблюдаемой проблемы или процесса (рис. 1). Для её построения необходимо

сначала выделить объект мониторинга, т.е. то, что собственно является контролируемой Проблемой. Далее нужно выделить несколько достаточно крупных факторов, которые непосредственно влияют на Проблему. Это так называемые элементы первого уровня. Затем каждый из полученных элементов при необходимости разбивается на несколько подфакторов (элементы второго уровня) и т.д. Несмотря на то, что эта стадия разработки систем информационного мониторинга является наименее формализованной, обычно она не вызывает трудностей у экспертов.

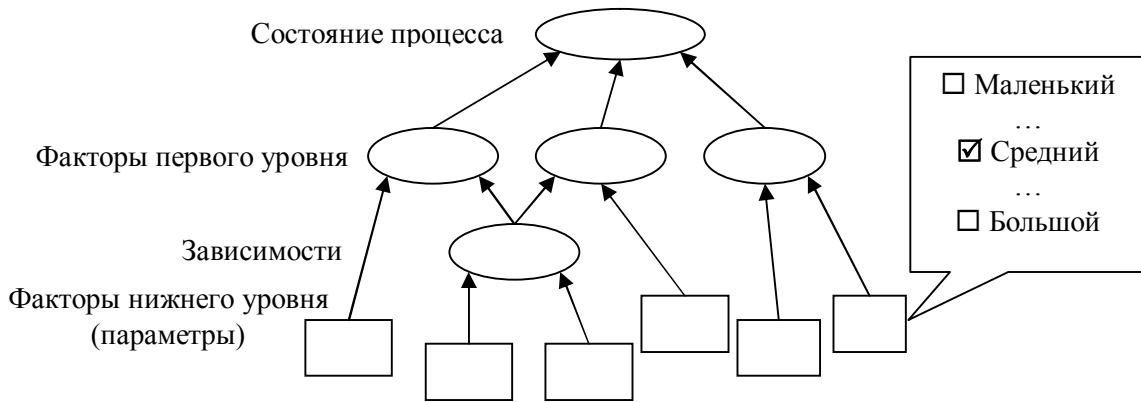
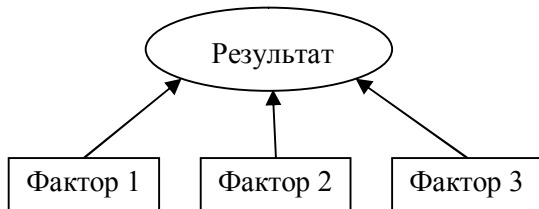


Рис. 1. Схема структурной модели проблемы

1.2. Создание поведенческой модели

Если структурная модель выявляет наличие зависимости между факторами предметной области, то поведенческая модель определяет характер этой зависимости (рис. 2). В моделях, например, физических процессов эти зависимости могут быть выражены формулами, дифференциальными уравнениями и т.д. Однако в других областях такое описание может оказаться неприемлемым в силу следующих причин:

- математической модели интересующей проблемы не существует;
- математическая модель слишком сложна;
- невозможно с достаточной точностью измерить все параметры модели.



$$\text{Результат} = F(\text{Фактор1}, \text{Фактор2}, \text{Фактор3})$$

Рис. 2.

Тем не менее, люди способны принимать верные решения в таких «непонятных» для машины ситуациях на основе так называемых «приближённых рассуждений». Одним из способов математической формализации такого способа принятия решений является теория нечётких множеств [1, 16].

Эксперт-разработчик и пользователи системы информационного мониторинга описывают состояния различных факторов предметной области и связи между ними в виде лингвистических значений и лингвистических правил соответственно. Субъективная степень удобства такого описания зависит от набора и состава таких лингвистических значений. Предполагается, что система также может интегрировать оценки различных экспертов. Поэтому желательно иметь гарантии того, что разные эксперты описывают один и тот же объект наиболее «единообразно». Таким образом, возникает проблема выбора оптимального множества значений для каждого фактора. Возможны два критерия оптимальности:

1. под оптимальными понимаются такие множества значений, при использовании которых человек испытывает минимальную неопределенность при описании объектов;

2. если объект описывается некоторым количеством экспертов, то под оптимальными понимаются такие множества значений, которые обеспечивают минимальную степень рассогласования описаний.

Эти вопросы подробно исследуются в [7], где показано, что мы можем сформулировать методику выбора оптимального множества значений качественных признаков. Более того, показано, что множества значений, оптимальные по критериям (1) и (2), совпадают.

Опыт разработки систем информационного мониторинга показал, что наиболее трудоемким этапом является построение именно поведенческой модели – модели правил. Трудности здесь объясняются тем, что специалистам не всегда удается сформулировать свои знания не только в виде четких формул, но даже в виде лингвистических правил. Причинами этого могут быть и недостаточная компетентность, и так называемые «интуитивные» знания, которые порой весьма трудно выразить вербально их носителям.

Одним из выходов из этой ситуации является технология «добычи знаний» (knowledge mining или data mining) [15]. Эта сравнительно молодая технология появилась в районе 90-х годов XX века и применяется для автоматического «извлечения» структурированных знаний, например, в виде лингвистических правил, из больших объемов данных, хранящихся в соответствующих базах данных. Эта же технология может использоваться для построения поведенческой модели нашей предметной области не только на основе знаний экспертов (которые они могут сформулировать), но и на основе исторических данных.

Важным отличием «добычи знаний» от статистических методов, которые также помогают находить закономерности в больших объемах данных, является то, что на выходе «добычи знаний» получаются понятные экспертам лингвистические правила, которые могут быть в дальнейшем ими уточнены и дополнены для улучшения адекватности модели. Таким образом, эта технология поможет заполнить тот пробел, который образовался вследствие того, что в силу различных причин экспертам не всегда удается сформулировать свои знания.

В работе [9] описан алгоритм направленного поиска ассоциативных правил, разработанный специально для применения в системах информационного мониторинга. Алгоритм был реализован программно, и, как показали практические испытания, работает достаточно быстро даже на персональных компьютерах среднего уровня.

Другим подходом к сокращению трудозатрат на разработку поведенческой модели является использование более высокоуровневых средств описания поведения системы. Вместо полного набора правил вида «Если параметр_1 = значение_1, ... параметр_n = значение_n, то результат = значение_результата», эксперт может описывать характер зависимости такими высказываниями, как, например, «при максимальных значениях всех аргументов значение результата максимально», «результат слабо изменяется при изменении второго аргумента», «при совместном возрастании второго и четвертого аргументов результат медленно улучшается» и т.п. Такой подход был впервые предложен в работе [6]. Дальнейшее его развитие приведено ниже в этой работе.

Также возможно совместное применение обеих технологий: характер зависимости может быть частично выяснен методом «добычи знаний» на основе исторических данных и дополнен описаниями эксперта, после чего метод восстановления зависимости по описанию скомбинирует имеющиеся знания в полную систему правил.

1.3. Возможности многократного использования модели

Может возникнуть ситуация, что система информационного мониторинга будет применяться для контроля состояния большого числа однотипных проблем, отличающихся составными частями: например, мониторинг проектов (см. пример ниже), мониторинг (экономического, экологического и т.п.) состояния региона, мониторинг безопасности объектов и т.д. Для таких случаев была предложена следующая схема разработки и использования системы:

1. *Разработка модели.* Эксперты составляют фрагменты модели, разрабатывают стандартные наборы правил для операторов агрегирования и правила выбора лингвистических значений.

2. *Адаптация модели.* Инженер (не обязательно являющийся экспертом в предметной области, но имеющий опыт в использовании технологии информационного мониторинга), используя подмодели, составляет модель для конкретного проекта.
3. *Инициализация и использование.* Пользователь (аналитик), используя правила выбора лингвистических значений, заполняет модель данными и получает необходимые результаты.

Графическое изображение описанной схемы приведено на рис. 3.

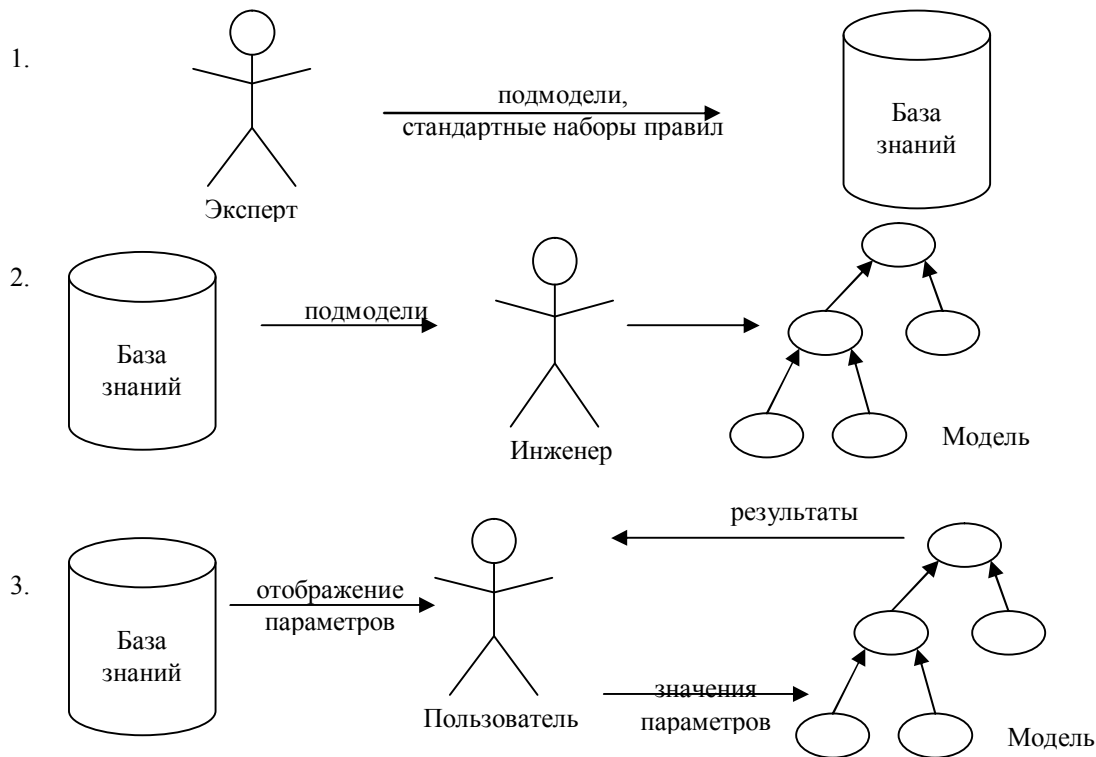


Рис. 3. Схема использования системы

1.4. Использование системы информационного мониторинга

Варианты использования системы:

- Оценка текущего состояния проблемы в целом или отдельных ее аспектов.
- Прогнозирование дальнейшего развития проблемы.
- Поиск оптимальных воздействий.

В отличие от остальных способов использования системы, последний вариант представляет собой обратную задачу («какими должны быть значения параметров, чтобы получился требуемый результат?»). Обычно такие задачи, с одной стороны, сложнее с вычислительной точки зрения, с другой стороны, представляют больший интерес для пользователя.

Более подробное исследование этого вопроса приведено ниже в этой работе.

2. Формализация и основные задачи

Теоретической моделью системы служит Схема Функциональных Элементов (СФЭ) [10]. Функциональные элементы вычисляют функции k -значной логики. Такая формализация полностью описывает случай дискретных систем; в нечётком случае моделируется только набор «если – то» правил, составляющий основу оператора агрегирования, а такие аспекты, как выбор функций принадлежности и метода нечёткого логического вывода, игнорируются.

СФЭ, используемые при моделировании систем информационного мониторинга, удовлетворяют следующим ограничениям:

- Схема обладает единственным выходом. Функцию, реализуемую схемой на этом выходе, будем обозначать $F(x_1, \dots, x_N)$.
- Число переменных у функциональных элементов ограничено сверху (максимальное число переменных обозначим n).

При практическом построении и использовании систем информационного мониторинга возможно использование неоднородных функций (k -«значность» логики различна для разных элементов в пределах одной схемы, см. например [2]). В этой работе мы будем считать k постоянным, т.к. это значительно упрощает обозначения, а переход к неоднородному случаю тривиален. Если это не оговорено отдельно, считаем, что $k \geq 3$.

2.1. Задача оптимального распределения ресурсов

Содержательная постановка

Данная задача возникает в ходе эксплуатации системы: в случае возможности влияния на некоторые параметры наблюдаемой проблемы, системы информационного мониторинга способны не только прогнозировать результат изменений значений параметров, но и находить оптимальный (с точки зрения соотношения вложенных средств и величины достигнутых изменений) вектор изменения параметров (решать обратную задачу).

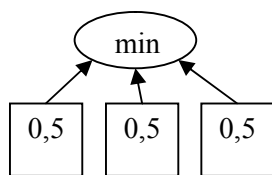


Рис. 4.

Ранние версии систем информационного мониторинга для этой цели использовали поиск «критического пути» – элемента модели, изменение значения которого (с сохранением значений всех остальных элементов) вызовет наибольшие изменения в состоянии всей проблемы. Однако во многих ситуациях такой подход не приводит к результату. Примером может служить схема, изображённая на рисунке: увеличение значения любого единственного параметра не окажет влияние на результат – необходимо увеличить значения всех факторов одновременно.

Предлагаемый нами подход лишён этого недостатка – в рассматриваемом дискретном случае он всегда позволяет найти оптимальное решение. Сохраняя связь с предыдущим подходом, будем называть множество факторов, совместное изменение которых оказывает наибольшее влияние на результат, «критическим множеством».

При таком обобщении необходимо также учесть различную природу ресурсов, требуемых для изменения значений параметров. Например, если в качестве ресурса выступают денежные средства, общая стоимость вектора управления будет вычисляться как сумма стоимостей его компонент; если же ресурсом является время и изменения могут осуществляться одновременно, общей стоимостью будет максимум стоимостей.

Формализация и труднорешаемость

Итак, формальная постановка задачи выглядит следующим образом:

$F(x_1, \dots, x_N)$ – функция k -значной логики, заданная схемой функциональных элементов над базисом, состоящим из всех функций от n и менее переменных; a_1, \dots, a_N – начальные значения переменных; $C_i(x)$ – стоимость присвоения i -ой переменной значения x (из текущего состояния). Для заданного C необходимо максимизировать $F(x_1, \dots, x_N)$ при ограничении $\perp_i C_i(x_i) \leq C$, где \perp – бинарная операция, удовлетворяющая аксиомам коммутативности, ассоциативности и монотонного убывания, которую мы будем называть функционалом стоимости.

В общем случае NP-полнота этой задачи очевидна для любого функционала стоимости. При фиксации же функционала стоимости могут возникать частные случаи, допускающие решение за полиномиальное время. С другой стороны, в некоторых случаях даже при внесении жёстких ограничений на базис функциональной схемы задача останется NP-полной. Эти ситуации мы продемонстри-

руем на примере двух наиболее часто встречающихся функционалов стоимости – суммы и максимума.

Теорема 1.1. Если $\perp(x, y) = \max(x, y)$, а базис СФЭ содержит только монотонные функции, то задача разрешима за линейное (от сложности схемы) время.

Теорема 1.2. Если $\perp(x, y) = x + y$, то задача оптимального распределения ресурсов NP-полна в следующей постановке:

$F(x_1, \dots, x_N)$ – функция k -значной логики ($k \geq 2$), заданная схемой функциональных элементов над множеством, содержащим функции от 2 переменных \min и \max . Для заданного натурального C необходимо проверить, существует ли такой набор $\alpha \in E_k^N, |\alpha| \leq C$, такой, что $F(\alpha) > 0$.

Теорема 1.3. Для СФЭ, граф которых является деревом, существует алгоритм, решающий задачу оптимального распределения ресурсов за время $C(n) \cdot |V|$, где $|V|$ – число вершин, $C(n)$ – величина, зависящая только от n – максимального числа переменных базисной функции.

Более подробно результаты этого раздела приведены в работе [3].

2.2. Синтез операторов агрегирования информации по экспертным описаниям

Задача выбора оператора агрегирования информации – определение функции, характеризующей зависимость некоторой величины от наблюдаемых параметров – возникает при разработке большинства систем сбора и обработки информации. Особую роль эта задача играет для систем информационного мониторинга. В работе [6] был впервые рассмотрен подход на основе нечётких условий: на множестве функций k -значной логики от n переменных на основе экспертных описаний задавалась функция принадлежности, и в качестве оператора агрегирования выбиралась функция k -значной логики, степень принадлежности которой максимальна. Однако в этих работах было описано всего два типа условий («значение в точке» и «локальное поведение по одной переменной»), и алгоритмы поиска функции были переборными.

Здесь мы предлагаем более общий подход к описанию функций k -значной логики нечёткими условиями. Мы покажем, как в этом подходе можно реализовать предыдущие результаты, докажем NP-полноту задачи выбора оператора по нечётким описаниям, а также приведём полиномиальные алгоритмы решения этой задачи в некоторых частных случаях.

Постановка задачи и формулировка результатов

Функции $E_K^n \rightarrow E_k$, где $E_l = \{0, 1, \dots, l-1\}$, будем называть функциями K, k -значной логики. Множество всех таких функций от n фиксированных переменных обозначим $P_{K,k}(n)$.

Нечётким условием на функции K, k -значной логики от n переменных мы называем отображение $\mu: P_{K,k}(n) \rightarrow [0; 1]$. Значение $\mu(f)$ называем степенью выполнения условия для функции $f \in P_{K,k}(n)$.

В нашей работе мы предлагаем задавать нечёткие условия с помощью управляющей системы, которую мы назовём *граф нечёткого условия*.

Граф нечёткого условия C – это тройка $\langle G, \rho, T \rangle$, где:

$G = \langle V, E \rangle$ – неориентированный граф без кратных рёбер с множеством вершин $V = \{(\alpha, a), \alpha \in E_K^n, a \in E_k\}$ и некоторым множеством рёбер E ;

$\rho: E \rightarrow [0; 1]$ – веса рёбер;

$T : [0;1]^2 \rightarrow [0;1]$ – Т-норма – бинарная операция, удовлетворяющая свойствам коммутативности, ассоциативности, монотонного неубывания и ограниченности: $T(0, x) = 0$, $T(1, x) = x$. (Т-нормы – обобщения операций «И» в нечёткой логике. Наиболее простыми примерами Т-норм являются функции минимума и произведения.

Граф нечёткого условия C реализует нечёткое условие $\mu : P_{K,k}(n) \rightarrow [0;1]$, степень выполнения которого для произвольной функции $f(x_1, \dots, x_n) \in P_{K,k}(n)$ вычисляется следующим образом:

- из множества вершин V графа G выделяется подмножество $V_f = \{(\alpha, f(\alpha)), \alpha \in E_K^n\}$;
- в графе G выделяется подграф $G_f = \langle V_f, E_f \rangle$, индуцированный подмножеством V_f ;
- значением $\mu_c(f)$ является значение Т-нормы от весов всех рёбер подграфа G_f , т.е.

$$\mu_c(f) = T_{e \in E_f} \rho(e). \text{ Если } E_f \text{ пусто, полагаем } \mu_c(f) = 1.$$

Класс всех графов нечёткого условия для фиксированных n, K, k и T обозначим $\Phi(K, k, n, T)$.

Степенью выполнимости нечёткого условия μ назовём величину $\mu^{\max} = \max_{f \in P_{K,k}(n)} \mu(f)$.

Задача определения степени выполнимости заключается в том, чтобы для данного нечёткого условия μ_c , заданного графом нечёткого условия C , и данного рационального $0 < \alpha \leq 1$ проверить, верно ли, что $\mu_c^{\max} \geq \alpha$.

Задача нахождения оптимальной функции заключается в нахождении для данного нечёткого условия μ_c , заданного графом нечёткого условия $C \in \Phi(K, k, n, T)$, хотя бы одной функции $f \in P_{K,k}(n)$, такой, что $\mu_c(f) = \mu_c^{\max}$.

Заметим, что решение задачи нахождения оптимальной функции позволяет легко найти решение задачи определения степени выполнимости. Обратное неверно.

Теорема 2.1. При $C \in \Phi(K, k, n, T)$ задача определения степени выполнимости нечёткого условия μ_c является NP-трудной для любого фиксированного $0 \leq \alpha \leq 1$, любого фиксированного $K \geq 2$ любого фиксированного $k \geq 3$ и любой фиксированной Т-нормы T .

Теорема 2.2. При справедливости гипотезы $P \neq NP$ не существует приближённого полиномиального алгоритма, решающего задачу определения степени выполнимости с результатом, отклоняющимся от истинного значения не более чем в фиксированное число раз.

Дальнейшие результаты работы посвящены выявлению случаев, разрешимых за полиномиальное время.

Теорема 2.3. Если $C \in \Phi(K, 2, n, \min)$, то задача нахождения оптимальной функции разрешима за время $O(K^{2n})$ при $n \rightarrow \infty$.

Замечание. Так как входом алгоритма является граф нечёткого условия, имеющий $2K^n$ вершин, то такая оценка сложности является полиномиальной.

Граф нечёткого условия назовём *параллельным*, если все рёбра в нём проводятся через пары вершин вида $(\alpha, a), (\alpha + \delta, b)$, где $\delta \in E_K^n$ – фиксированный вектор. Класс всех параллельных графов нечёткого условия обозначим $\Pi(K, k, n, T)$.

Теорема 2.4. Если $C \in \Pi(K, k, n, T)$, то задача нахождения оптимальной функции разрешима за время $O(K^{n+k+1})$ при $n \rightarrow \infty, K \rightarrow \infty$.

Граф нечёткого условия назовём *локальным*, если в нём ребро между вершинами (α, a) и (β, b) может быть проведено, только если $\max_i |\alpha_i - \beta_i| = 1$. Класс всех локальных графов нечёткого условия обозначим $\Lambda(K, k, n, T)$

Теорема 2.5. Если $C \in \Lambda(K, k, n, T)$, то при $n = 1$ задача нахождения оптимальной функции разрешима за время $O(k^2 \cdot K)$ при $K \rightarrow \infty, k \rightarrow \infty$, а для любого фиксированного $n \geq 2$ задача определения степени выполнимости условия является NP-трудной.

Примеры условий

В этом разделе мы приведём некоторые наиболее распространённые типы нечётких условий, в том числе и описанные в работе [6], которые можно реализовать в нашем подходе.

Условие на значение в точке. Это условие определяет желаемое значение функции в некоторой точке вне зависимости от принимаемых ей значений в других точках. Примерами условий такого типа могут быть: «при максимальных значениях всех аргументов значение функции максимально», «значение функции не превосходит значения третьего аргумента» (это условие представляет собой конъюнкцию нескольких точечных условий).

Локальное условие по одной переменной. Это условие определяет желаемое поведение функции на паре наборов, отличающихся на единицу по одной выбранной переменной. Условия такого типа могут задаваться высказываниями: «функция не убывает по первой переменной», «функция слабо изменяется при изменении второго аргумента» и т.п.

Локальное условие по нескольким переменным. Это условие определяет желаемое поведение функции на наборах, отличающихся на единицу по нескольким выбранным переменным. Условия такого типа могут задаваться высказываниями: «при совместном возрастании второй и четвёртой переменных функция слабо возрастает» и т.п.

Логические операции. Мы можем комбинировать условия с помощью логических операций «И» и «ИЛИ».

Более подробно результаты этого раздела приведены в работе [4].

3. Примеры создания систем информационного мониторинга

В этом разделе приводятся несколько примеров создания и применения систем информационного мониторинга в различных предметных областях. Описываются основные этапы разработки и способы использования, приведённые в разделе 1.2, а также демонстрируются различные подходы, диктуемые особенностями задач (масштабом контролируемой ситуации, особенностями используемых данных и т.д.).

3.1. Система мониторинга ядерных технологий

Система мониторинга ядерных технологий проекта DISNA (Development of an Intelligent System for Monitoring and Evaluation of Peaceful Nuclear Activities) разрабатывалась по заказу Международного Агентства по Атомной Энергии (МАГАТЭ). Цель системы – помощь сотрудникам управления гарантий Агентства в выявлении недеklarированной ядерной активности стран-участниц. Учитывая специфичность этой деятельности Агентства, приведем краткое описание таких механизмов.

Механизмы реализации международных гарантий МАГАТЭ

Одна из главных целей МАГАТЭ – предоставление надежных аргументированных заключений мировому сообществу о соблюдении странами-участницами взятых на себя обязательств в ядерной области. Для решения этой задачи в Агентстве существует специальная структура – управление гарантий (department of Safeguards). Техническая цель гарантий – своевременное обнаружение несоот-

ветствия объявленных мирных ядерных действий и существующих запасов ядерного материала и технологий, делающих возможным изготовление ядерного оружия или других ядерных взрывчатых устройств. Такие необъявленные действия могут касаться собственно ядерного материала, технологий его добычи и обработки, материальных средств и других ресурсов.

Ясно, что для углубленного анализа деятельности страны эксперты Агентства должны анализировать не только официальную декларацию страны, но и проводить детальный анализ ее экономики, научного, технологического и кадрового потенциала. Для этого они должны обладать средствами доступа к информации о стране, анализа и моделирования информационных потоков различных сторон деятельности страны.

На основе всей доступной информации эксперт Агентства делает заключение о состоянии ее ядерной программы. Обычно в заключении содержатся ответы на вопросы:

- Является ли официальная ядерная программа страны взвешенной и последовательной?
- Находятся ли реальные действия страны в ядерной области в соответствии с декларированными (официальными)? Подтверждается ли это анализом образцов окружающей среды (почва, вода, воздух)?
- Находятся ли в соответствии производство, импорт, накопленные запасы ядерных материалов и объявленная в декларации официальная ядерная программа страны?
- Совместим ли реальный импорт технологий, оборудования, специфических неядерных материалов и объявленная в декларации официальная ядерная программа страны?

Ответы на некоторые сформулированные вопросы можно получать, используя различные методы наблюдений за состоянием окружающей среды, различные методы анализа информации, и т.п., но для получения полной картины нужна специальная система слежения за развитием ядерной программы страны и ее оценки.

Для создания такой системы МАГАТЭ объявило международный конкурс. В конкурсе приняли участие коллективы Германии, Финляндии, Австрии, России и представители других государств. Поддержку международного жюри экспертов в области международных гарантий и информационных технологий получило предложение Московского государственного университета им. М.В. Ломоносова, базирующееся на технологии информационного мониторинга.

Модель проблемы

Основу системы составила так называемая физическая модель (Physical Model), которая была разработана экспертами Агентства и стран-участниц в рамках подготовки усиленной системы гарантий и была доработана для использования в системе информационного мониторинга и оценки ядерной активности стран DISNA. Эта модель является внутренним документом Агентства и составляет несколько тысяч листов текста. Ее идея заключается в том, что в природе не существует ядерного материала, пригодного для изготовления оружия – его получение есть цепочка преобразований исходной руды; каждый шаг может включать несколько процессов, при этом выбор процесса для конкретного шага зависит от процессов, реализованных на предшествующих шагах.

В физической модели систематизированы, описаны и охарактеризованы все известные процессы для выполнения каждого шага, необходимого для производства оружейного материала. Таким образом, любой возможный маршрут производства от руды до оружейного материала описывается так некоторая комбинация процессов, идентифицированных, описанных и охарактеризованных в физической модели.

Инициализация и использование

Информационные источники Агентства могут быть разделены на три группы:

1. Информация, предоставляемая Агентству страной (отчет страны, протокольная информация, уведомления).

2. Информация, собираемая Агентством независимо (замеры состояния окружающей среды, инспекторские отчеты, информация третьих сторон – других международных организаций, спецслужб и т.п.).
3. Информация из открытых источников (базы данных Monterey, Института атомной энергии им. Курчатова, агентства Рейтер, FBIS, NNN, ИНТЕРНЕТ, источники на бумажных носителях – специализированные обзоры, журналы, газеты и т.д.).

Первая группа источников служит для получения информации об официальной деятельности страны в ядерной области и проверяется информацией из источников второй и третьей группы. Информация из источников первой группы предоставляется экспертам Агентства для оценки и анализа в соответствии с обязательствами страны и хранится в Агентстве. Объем такой информации сравнительно небольшой, все сообщения относятся к предметной области и должны учитываться. Источники таких сообщений считаются надежными по определению.

Сообщения из источников второй группы содержат информацию в основном среднего и высокого уровня надежности. Объем такой информации по сравнению с первой группой значительно больше и, следовательно, требует больших затрат на отбор и обработку. Не все сообщения из этой группы будут относиться к интересующей части предметной области и должны учитываться при анализе ее состояния.

Важной особенностью работы экспертов МАГАТЭ является возможность инспекционных поездок в страну и анализа ситуации «на месте». Поэтому процесс оценки ядерной программы страны носит итеративный характер, в процессе которого «подозрительные» моменты в поведении страны могут быть уточнены и подтверждены или опровергнуты. Планирование инспекции – это отдельная задача, которая как раз и базируется на углубленном анализе всей доступной информации.

Результаты

Система обеспечила решение следующих задач:

- предоставить инструмент для непрерывного мониторинга состояния ядерной программы страны;
- проводить оценку влияния полученной информации на состояние элементов модели (изменение или подтверждение их состояния);
- предоставить возможность исследовать состояние ядерной деятельности страны с несколькими уровнями детализации;
- обнаруживать несоответствие между объявленными в декларации возможностями страны по обработке ядерного материала и реальными возможностями страны, получаемыми Агентством на основе анализа информации из других доступных МАГАТЭ источников;
- оценивать важность любого обнаруженного несоответствия с точки зрения возможности производства страной оружейного урана/плутония;
- находить «критические пути» – важные с точки зрения производства оружейного ядерного материала элементы модели, текущая информация о которых является критической для обнаружения несоответствия между декларацией страны и ее реальными возможностями, установленными Агентством;
- обеспечить эксперта инструментом для моделирования возможных изменений состояния элементов модели и оценки влияния этих изменений на общее состояние ядерной программы страны.

Более подробное описание этого проекта приведено в работе [14].

3.2. Оценка и мониторинг проектов разработки высокотехнологичных изделий

Одной из первых и важных проблем любого инновационного проекта является проблема оценки способности коллектива разработчиков выполнить проект в необходимое время с требуемым качеством при заданном ресурсном обеспечении [12]. Примерами вопросов, которые возникают на этой начальной стадии, являются:

- Достаточно ли у коллектива разработчиков навыков и ресурсов для завершения проекта?

- Какие части проекта представляют наибольшие трудности?
- Каким образом различные вложения средств (например, обучение, закупка оборудования, программного обеспечения) повлияют на возможности коллектива?
- Как оптимально улучшить возможности коллектива разработчиков?

В настоящее время на эти вопросы отвечают эксперты, менеджеры проекта, основываясь на своём личном опыте и интуиции. Цена ошибки на этом этапе – увеличение стоимости проекта, увеличение времени разработки изделия – вплоть до провала проекта. Поэтому любое повышение надёжности оценок и ответов на сформулированные вопросы является важным, иногда критичным, для любого серьёзного проекта.

Для повышения надёжности, объективности прогнозов экспертов и снижения материальных и временных затрат на этом этапе разработки была создана система оценки и мониторинга проектов на примере разработки изделий микроэлектроники. В рамках разработки системы происходит формализация знаний экспертов и автоматизация процессов обработки информации о проектном коллективе для получения ответов на приведённые выше и другие вопросы. Отметим, что данная предметная область (микроэлектроника) не обладает какими-либо специфическими свойствами, удобными для разработки системы. Ее выбор – лишь удачное стечение обстоятельств: Cadence Design Systems (www.cadence.com) – лидер рынка САПР микроэлектроники – имеет необходимый опыт (модель «виртуальной САПР» Cadence VCAD [13]) и потребность в такой системе как инструменте совершенствования услуг VCAD. Описываемый ниже прототип разработан с привлечением экспертов.

Создание структурной модели

Для разработанного приложения экспертами были выделены следующие факторы, определяющие способности проектного коллектива (рис. 5):

1 уровень: способность завершить отдельные стадии процесса разработки (разработка, тестирование, отладка и т.д.);

2 уровень:

- Навыки – различные для каждой стадии;
- Оборудование – различное для каждой стадии (вычислительные ресурсы, программное обеспечение);
- Инфраструктура – общая для всех стадий (доступ к литературе, работа сети и т.д.).

Дальнейшие уровни иерархии содержат атрибуты, более детально описывающие факторы второго уровня. Модель содержит более 60 элементов и более 70 связей, объединенных в 5 уровней иерархии.

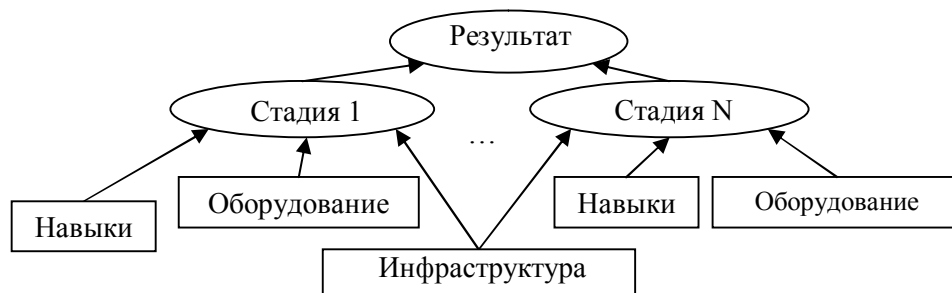


Рис. 5. Структура модели системы

Создание поведенческой модели

В этом проекте ввод лингвистических правил осуществлялся вручную – эксперт не испытывал затруднений в указании значения для любой комбинации значений параметров. Для сокращения трудозатрат использовались следующие особенности предметной области:

- Для большинства элементов модели минимальные (максимальные) значения факторов-параметров влекли за собой минимальное (соответственно, максимальное) значение результата.
- В большинстве случаев набор правил обладал свойством монотонности: бóльшие значения факторов-параметров влекли за собой бóльшее значение результата.
- Многие факторы обладали свойством критичности: значение результата не могло превышать значение этого фактора.
- Для многих элементов модели элементу было проще указать весовые коэффициенты, определяющие степень влияния параметров на результат. На основе этих коэффициентов строилась первичная система правил, которая впоследствии корректировалась экспертом.

Инициализация

Для этого случая стадия инициализации очень проста – пользователю (менеджеру проекта) доступны все сведения о проектном коллективе, и ему лишь остаётся ввести их в систему.

Результаты

На момент написания этой работы функциональный прототип системы успешно прошёл первичное тестирование и перешёл в стадию опытной эксплуатации.

Более подробное описание этого проекта приведено в работе [5].

3.4. Возможные области применения технологии

Приведённые примеры показывают, что спектр применения технологии информационного мониторинга чрезвычайно широк. Необходимыми условиями применения технологии являются лишь иерархическая структура контролируемой проблемы и опыт решения задачи на экспертном уровне.

По мнению авторов, технологию также можно применить для решения следующих задач:

- Мониторинг систем безопасности различного рода;
- Мониторинг экономических и социальных государственных программ;
- Мониторинг деятельности предприятий (интересным представляется синтез технологии информационного мониторинга и Сбалансированной Системы Показателей [11]) и ведомств.

Заключение

В работе были описаны основные этапы практической разработки систем информационного мониторинга. Также была предложена формализация работы системы с использованием классического аппарата дискретной математики – схем функциональных элементов и функций k -значной логики, и в этой формализации описаны решения двух основных теоретических задач технологии: задачи оптимального распределения ресурсов и синтеза операторов агрегирования. Помимо этого, были приведены несколько примеров создания и применения систем информационного мониторинга в таких предметных областях как международная безопасность и менеджмент.

Литература

1. Заде Л.А. Понятие лингвистической переменной и его применение к принятию приближенных решений. – М.: Мир, 1976. – С. 165.
2. Кудрявцев В.Б. Функциональные Системы. – М.: Издательство МГУ, 1982. – С. 158.
3. Лебедев А.А. О задачах оптимального распределения ресурсов и проверки устойчивости для схем функциональных элементов в k -значной логике (готовится к печати).
4. Лебедев А.А. Синтез операторов агрегирования информации по экспертным описаниям (готовится к печати).
5. Лебедев А.А., Рыжов А.П. Оценка и мониторинг проектов разработки высокотехнологичных изделий на примере микроэлектроники // Интеллектуальные системы. – 2008. – Том 11. – Вып. 1-4. – С. 55-82.
6. Рыжов А.П. Об агрегировании информации в нечетких иерархических системах // Интеллектуальные системы. – 2001. – Том 6. – Вып. 1-4. – С. 341-364.
7. Рыжов А.П. О степени нечеткости размытых характеристик // Математическая кибернетика и ее приложения в биологии. Под редакцией Л.В. Крушинского, С.В. Яблонского, О.Б. Лупанова. – М.: Издательство МГУ, 1987. – С. 60-77.
8. Рыжов А.П. Оценка степени нечеткости и ее применение в системах искусственного интеллекта // Интеллектуальные системы. – М.: МНЦ КИТ, 1996. – Т.1. – Вып. 1-4. – С. 95-102.
9. Рыжов А.П., Расторгуев В.В. Методы извлечения нечетких ассоциативных правил в системах информационного мониторинга // Труды международных научно-технических конференций «Интеллектуальные системы» и «Интеллектуальные САПР», 3-10 сентября 2006 г., Дивноморское, Россия. – М.: Физматлит, 2006. – Том 1. – С. 70-81.
10. Яблонский С.В. Основные понятия кибернетики // Проблемы кибернетики. – 1959. – Вып. 2. – С. 7-38.
11. Kaplan R.S., Norton D.P. The balanced scorecard: measures that drive performance // Harvard Business Review Jan–Feb 1992. – Pp. 71-80.
12. Martin Michael J.C. Managing innovation and entrepreneurship in technology based firms. NY: John Wiley & Sons, Inc., 1994. – P. 402.
13. Matzke W.E., Strube G., Schmidt-Habich H., Drenan L. VCAD – a virtual enterprise collaboration model impacting the semiconductor industry // IASTED International Conference on Knowledge Sharing & Collaborative Engineering (KSCE 2004).
14. Ryjov A., Belenki A., Hooper R., Pouchkarev V., Fattah A., Zadeh L.A. Development of an Intelligent System for Monitoring and Evaluation of Peaceful Nuclear Activities (DISNA) // IAEA, STR-310. Vienna, 1998. – P. 122.
15. Usama M. Fayyad (Ed.). Advances in Knowledge Discovery and Data Mining. MIT Press, 1996. – P. 560.
16. Zadeh L.A. Fuzzy sets // Information and Control. – 1965. – Vol.8. – Pp. 338-353.