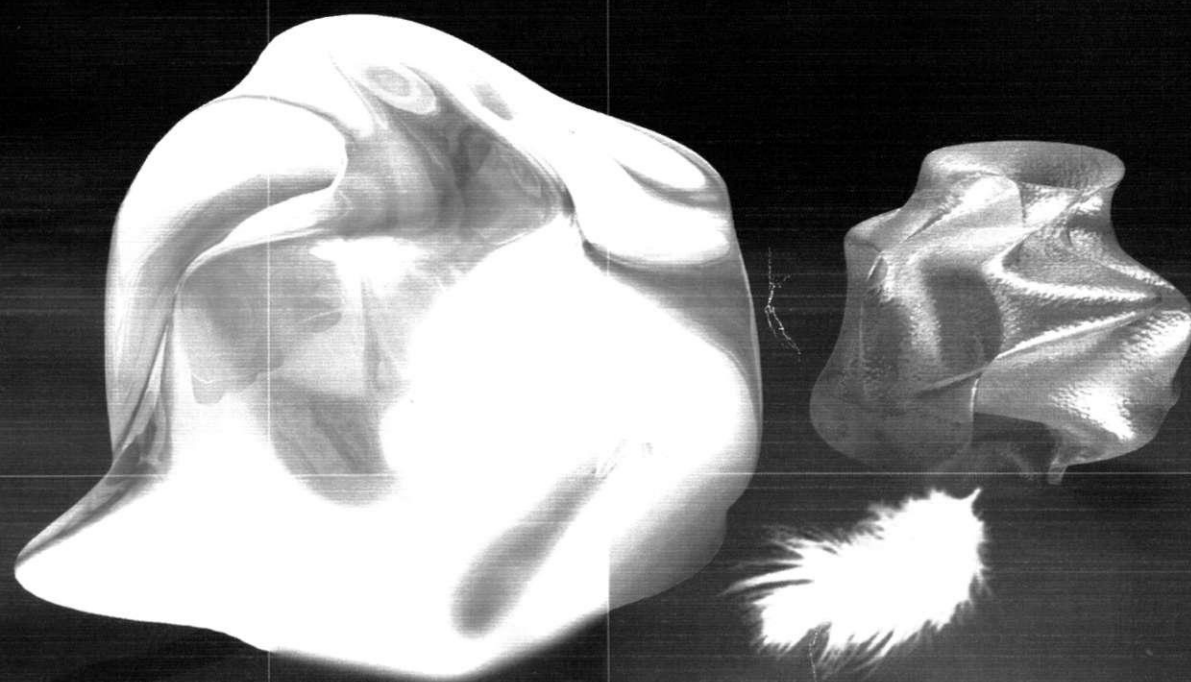


Том 3, № 4, Декабрь 2008

ISSN 1819-4362

# НЕЧЕТКИЕ СИСТЕМЫ И МЯГКИЕ ВЫЧИСЛЕНИЯ



Научный журнал Российской ассоциации  
нечетких систем и мягких вычислений

**ИНТЕЛЛЕКТУАЛЬНЫЕ ВЫЧИСЛЕНИЯ В СИСТЕМАХ  
ИНФОРМАЦИОННОГО МОНИТОРИНГА****Лебедев А.А., Рыжов А.П.**Кафедра математической теории интеллектуальных систем,  
механико-математический факультет, МГУ им. М.В.Ломоносова, Москва

---

*Поступила в редакцию 25.09.2008, после переработки 04.10.2008.*

---

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, а также приводятся теоретические результаты, гарантирующие оптимальность их работы. Помимо этого приводятся несколько примеров создания и применения систем информационного мониторинга в различных предметных областях с описанием основных этапов разработки и способов использования и демонстрации различных подходов, диктуемых особенностями задач (масштабом контролируемой ситуации, особенностями используемых данных и т.д.).

Information Monitoring Technology was developed for analysing complex systems or processes that do not have an appropriate mathematical model. Tools based on this technology (called information monitoring systems) can utilise all available information to provide estimates of the development of the monitored system/ process and find optimal strategies for controlling its behaviour. In this paper, we discuss the principal aspects of developing such systems and provide some theoretical results that guarantee their correctness and optimality. We also give several examples of developing and applying information monitoring systems for various application domains. We describe the main development stages as well as various ways of usage, and demonstrate some differences in the approaches caused by differences in the addressed problems' characteristics.

**Ключевые слова:** информационный мониторинг, нечеткие иерархические системы.

**Keywords:** information monitoring, fuzzy hierarchic systems.

**1. Введение**

Содержание работы многих классов специалистов в современном информационном обществе можно представить как работу по определенным 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 проблемы исходя из целей, стоящих перед специалистом. Можно привести следующие примеры проблем из различных областей:

- моделирование поведения клиента (маркетинг);
- сегментирование рынков (маркетинг);
- продвижение кандидата на выборах (политология);
- 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 информации, т.е. фиксация информации в виде статей, газетных заметок, компьютерном виде, аудио - и видеoinформация и т.п.;
- фрагментарность: информация чаще всего относится к какому-либо фрагменту проблемы, причем разные фрагменты могут быть по-разному «покрыты» информацией;
- разноуровневость: информация может относиться ко всей проблеме в целом, к некоторой ее части, к конкретному элементу проблемы;
- различная степень надежности: информация может содержать конкретные данные различной степени надежности, косвенные данные, результаты выводов на основе надежной информации или косвенные выводы;
- возможная противоречивость: данные из различных источников могут совпадать, слегка различаться или вообще противоречить друг другу.
- изменяемость во времени: проблема/процесс развивается во времени, поэтому и информация в разные моменты времени об одном и том же элементе проблемы может и должна различаться.

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 (отвечать на вопросы типа «Что будет, если ... ?»). 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 имеет большое практическое значение и 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.

## **2. Основные этапы разработки и использования систем информационного мониторинга**

### *2.1 Предпосылки применения технологии информационного мониторинга*

Применение технологии информационного мониторинга наиболее эффективно (по сравнению с другими автоматизированными средствами) для контроля и управления состоянием проблем или процессов, обладающих следующими свойствами:

- Отсутствие (или невозможность применения) адекватной математической модели.
- Иерархическая структура проблемы.
- Наличие большого числа разнородных факторов.
- Наличие экспертов предметной области.

### *2.2 Этапы разработки системы*

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.

### *2.3 Создание структурной модели*

Структурная модель представляет собой иерархически организованный набор факторов, определяющих состояние наблюдаемой проблемы или процесса (рис.

1). Для ее 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. При этом нужно помнить, что для функционирования системы данные о состоянии всех факторов модели должны быть доступны, в модель следует включать только те факторы, по которым можно легко и достаточно дешево получить требуемую информацию, – в противном случае избыток факторов в модели повлечет существенное увеличение стоимости использования технологии. Таким образом, при построении структурной модели необходимо руководствоваться следующими двумя (противоречивыми) критериями:

1. *Полнота.* Это означает, что в модель должны быть включены все факторы, имеющие важность для описания предметной области.
2. *Доступность информации.*



Рис. 1: Схема структурной модели проблемы

Несмотря на то, что эта стадия разработки систем информационного мониторинга является наименее формализованной, обычно она не вызывает трудностей у экспертов.

#### 2.4 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;

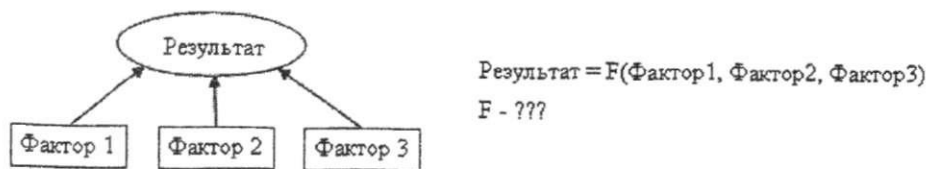


Рис. 2: Функция агрегирования

- Невозможно с достаточной точностью измерить все параметры модели.

Тем не менее, люди способны принимать верные решения в таких «непонятных» для машины ситуациях на основе так называемых «приближенных рассуждений». Математическая формализация такого способа принятия решений была впервые предложена в работах Заде [2, 18]. Последующие его работы положили начало новому продуктивному разделу математики, теории управления и искусственного интеллекта – теории нечетких множеств.

Идея заключается в том, что множество возможных значений каждого фактора можно представить в виде набора «плавно перетекающих друг в друга» лингвистических значений (например, вместо некоторой численной шкалы, параметру качество мы можем присваивать, скажем, значения «очень высокое», «высокое», «удовлетворительное», «неудовлетворительное» и т.д.). После этого зависимости можно описывать в виде набора лингвистических правил, связывающих лингвистические значения параметров с лингвистическим значением результата (например, «если температура высокая и пульс очень частый, то состояние пациента тяжелое»). Именно в таком виде описывается поведенческая модель в системах информационного мониторинга. Это позволяет единообразно работать с информацией различного рода, уровня общности и степени достоверности.

Итак, 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. Это значит, что один объект может описываться разными экспертами. Поэтому желательно иметь гарантии того, что разные эксперты описывают один и тот же объект наиболее «единообразно». Исходя из сделанных замечаний, мы можем сформулировать первую проблему следующим образом: можно ли, учитывая некоторые особенности восприятия человеком объектов реального мира и их описания, сформулировать правило выбора оптимального множества значений признаков, по которым описываются эти объекты? 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 [6]. 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. Для решения этой проблемы могут привлекаться так называемые когнитологи или специалисты по знаниям: они пытаются структурировать знания экспертов о предметной области и таким образом построить хотя бы приближенную ее модель. Большим недостатком такого подхода является то, что задача структурирования знаний экспертов – это скорее искусство, чем технология, и вследствие этого привлечение подобных специалистов существенно удорожает процесс построения системы информационного мониторинга. Кроме того, когнитологи ничем не могут помочь в случае недостаточной компетентности экспертов.

Одним из выходов из этой ситуации является технология «добычи знаний» (knowledge mining или data mining) [17]. 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» получают 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 [8] 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 предложен в работе [5]. 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.

### *2.5 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.

### *2.6 Использование системы информационного мониторинга*

Доступны следующие варианты использования системы:

- *Оценка текущего состояния проблемы в целом или отдельных ее аспектов.*
- *Прогнозирование дальнейшего развития проблемы.* Пользователь системы, задавшись вопросом «а что будет, если этот фактор изменится некоторым образом?», может достаточно легко получить на него ответ.
- *Моделирование последствий действий.* Под действием понимается некоторое мероприятие, приводящее к одновременному изменению сразу нескольких факторов.
- *Поиск оптимальных воздействий.*

Остановимся подробнее на последнем варианте использования. В случае возможности влияния на некоторые параметры наблюдаемой проблемы или процесса, системы информационного мониторинга способны находить оптимальный (с точки зрения соотношения вложенных средств и величины достигнутых изменений) набор величин изменения параметров – так называемый вектор управления.



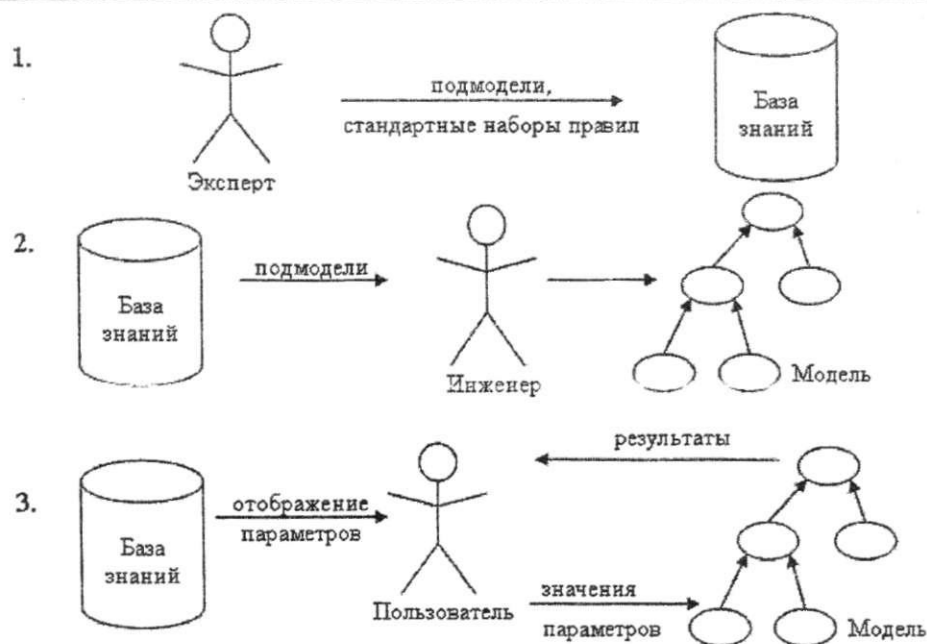


Рис. 3: Схема использования системы

В отличие от остальных способов использования системы, являющихся прямыми задачами (отвечающими на вопрос «каким будет значение результата при данных значениях параметров?»), в этом случае решается обратная задача («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.

Подробное исследование этого вопроса приведено ниже в этой работе.

### 3. Формализация и основные задачи

Теоретической моделью системы служит Схема Функциональных Элементов (СФЭ) [9]. Функциональные элементы схемы вычисляют функции  $k$ -значной логики. Такая формализация полностью описывает случай дискретных систем; в нечетком случае моделируется только набор «если - то» правил, составляющий основу оператора агрегирования, а такие аспекты, как выбор функций принадлежности и метода нечеткого 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  $n$ )

### 3.1 Задача оптимального распределения ресурсов

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

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, изображенная на рисунке: увеличение значения любого единственного параметра не окажет влияния на результат – необходимо увеличить значения всех факторов одновременно.

Предлагаемый нами подход лишен 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.

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  $\bigoplus_i C_i(x_i) \leq C$ , где  $\bigoplus$  – бинарная операция, удовлетворяющая аксиомам коммутативности, ассоциативности и монотонного неубывания, которую мы будем называть функционалом стоимости.

В общем случае NP-полнота этой задачи очевидна для любого функционала стоимости, т.к. NP-полна даже задача выполнимости – задача о существовании набора значений аргументов (вне зависимости от его стоимости), при котором функция принимает значение, отличное от исходного. При фиксации же функционала стоимости могут возникать частные случаи, допускающие решение за полиномиальное время. С другой стороны, в некоторых случаях даже при внесении жестких ограничений на базис функциональной схемы задача останется NP-полной. Эти ситуации мы продемонстрируем на примере двух наиболее часто встречающихся функционалах стоимости – сумме и максимуме.

**Утверждение 1.** Если  $\bigoplus(x, y) = \max(x, y)$  а базис СФЭ содержит только монотонные функции, то задача разрешима за линейное (от сложности схемы) время.

*Доказательство.* Алгоритм, решающий задачу в этом случае, выглядит следующим образом:

1. Для каждой переменной  $x_i$  находится максимальное значение  $a_i$ , такое что  $C_i(a_i) \leq C$ .
2. Вычисляется значение  $F(a_1, \dots, a_n)$ . В силу монотонности  $F$  оно и является искомым максимумом.

Число шагов алгоритма на шаге 1 пропорционально числу переменных, которое не превосходит сложность схемы. На шаге 2 число шагов пропорционально сложности схемы. Утверждение доказано.

**Теорема 1.** Если  $\perp(x, y) = x + y$ , то задача оптимального распределения ресурсов NP-полна в следующей постановке:

$F(x_1, \dots, x_n)$  – функция алгебры логики ( $k = 2$ ), заданная схемой функциональных элементов над множеством  $\{\&, \vee\}$  (отсюда автоматически следует, что  $F(0, \dots, 0) = 0$ ). Для заданного натурального  $C$  необходимо проверить, существует ли такой набор  $\alpha \in E_2^n, |\alpha| \leq C$ , такой что  $F(\alpha) = 1$ . (по определению,  $|\alpha| = \sum_{i=1}^n \alpha_i$ ).

*Доказательство.* Принадлежность задачи классу NP очевидна: сертификатом будет служить набор  $\alpha$ , удовлетворяющий условиям задачи. Для доказательства NP-трудности нашей задачи сведем к ней классическую NP-полную задачу МИНИМАЛЬНОЕ ПОКРЫТИЕ [12]. Нам дано множество элементов  $Y = \{\hat{y}_1, \dots, \hat{y}_M\}$ ,  $X = \{\hat{x}_1, \dots, \hat{x}_N\}$  – некоторое множество его подмножеств, и натуральное  $C$ . Необходимо проверить, можно ли выбрать  $l \leq C$  элементов  $\hat{x}_{i_1}, \dots, \hat{x}_{i_l} \in X, \bigcup_j \hat{x}_{i_j} = Y$ .

Данным  $Y$  и  $X$  сопоставим следующую СФЭ (рис. 4):

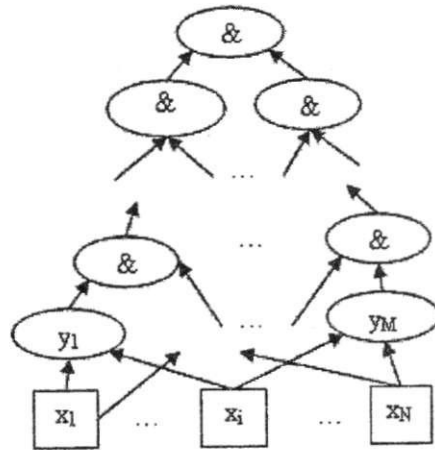


Рис. 4: СФЭ, решающая задачу о минимальном покрытии

Вплоть до предпоследнего слоя схема представляет собой бинарное дерево, все функциональные элементы вычисляют конъюнкции двух переменных. Элементы последнего слоя  $y_1, \dots, y_M$  вычисляют дизъюнкции некоторого числа переменных. Переменная  $x_i$  соединена с элементом  $y_j$  тогда и только тогда, когда  $\hat{x}_i \in \hat{y}_j$ .

Функцию, реализуемую этой СФЭ, обозначим  $F(x_1, \dots, x_N)$ . Нетрудно проверить, что нахождение вектора  $\alpha$ , удовлетворяющего условиям задачи, решает и

задачу о наименьшем покрытии (искомые индексы  $i_1, \dots, i_l$  – номера его ненулевых компонент). Теорема доказана.

**Следствие 1.** Задача 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  $F(\alpha) > 0$ .

**Следствие 2.** В формулировке Следствия 1 задача NP-полна и для функционала стоимости  $\perp(x, y) = xy$ .

Труднорешаемость задачи в общем случае приводит нас к необходимости выделения частных случаев, допускающих решение эффективными алгоритмами. Некоторые такие случаи были описаны в предыдущем разделе. Здесь мы опишем один случай, не зависящий от выбора функционала стоимости.

В случае, когда граф СФЭ является деревом, задача полиномиально разрешима при любом функционале стоимости даже для базиса, состоящего из всех функций от  $n$  и менее переменных. Более того, получен алгоритм, работающий за линейное (от сложности схемы) время.

**Алгоритм 1.** Дано:  $G = (V, E)$  – ориентированное дерево,  $L = \{l_1, \dots, l_N\}$  – множество листьев,  $\Phi : V$

$L \rightarrow P_k$  – операторы агрегирования, соответствующие неконцевым вершинам,  $\Psi : L \times E_k \rightarrow R^+$  – «стоимости» изменения значений переменных.

1. Разобьем вершины графа на уровни  $V = V_0 \cup V_1 \cup \dots \cup V_h$  в соответствии с длиной (единственного) пути до корневой вершины.
2. Двигаясь по уровням  $i$  от  $h$  до 1, доопределим  $\Phi$  на вершины из  $V \setminus L$  следующим образом:

Рассмотрим все вершины из  $V_i \setminus L$  – вершины уровня  $i$ , которые не являются листьями. Для каждой такой вершины  $p$  рассмотрим  $V_p = \{v_1, \dots, v_{n_p}\} \subset V_{i+1}$  – множество всех его непосредственных потомков.

Пусть  $\Phi(p) = f_p(u_1, \dots, u_{n_p}) \in P_k(n_p)$ . Тогда для всех  $y \in E_k$  вычисляем

$$\Psi(p, y) = \min_{\substack{(\alpha_1, \dots, \alpha_{n_p}): \\ f_p(\alpha_1, \dots, \alpha_{n_p}) = y}} \left\{ \prod_{j=1}^{n_p} \Psi(v_j, \alpha_j) \right\}.$$

3. На выходе алгоритма получаем функцию  $\hat{\Psi}(y) = \Psi(s, y)$  – минимальные стоимости присвоения выходной функции значения  $y$ .

**Утверждение 2** (время работы алгоритма 1). Время работы алгоритма не превосходит  $C(n)|V|$ .

**Доказательство.** Время работы шага 1 алгоритма –  $C_1|V|$  (очевидно). На шаге 2 каждая неконцевая вершина просматривается один раз, причем время, тратящееся алгоритмом на каждую вершину, ограничено величиной, зависящей только от  $n$ .

**Замечание.** Подобный метод можно применить и для графов более общего вида. Соответствующая работа готовится к публикации.

### 3.2 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 [5] был впервые рассмотрен подход на основе нечетких 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 нечеткими условиями. Мы покажем, как в этом подходе можно реализовать предыдущие результаты, докажем NP-полноту задачи выбора оператора по нечетким описаниям, а также приведем полиномиальные алгоритмы решения этой задачи в некоторых частных случаях.

Итак, мы хотим описать нечеткое условие на множестве функций  $k$ -значной логики от  $n$  переменных, т.е. определить отображение  $\mu : P_k(n) \rightarrow [0; 1]$ . Для этого мы введем конструкцию *граф нечеткого условия*.

Рассмотрим граф  $G$  следующего вида:

- Множество его вершин есть  $\{(\alpha, a), \alpha \in E_k^n, a \in E_k\}$ ;
- Ребра неориентированные;
- Допускаются петли и висячие вершины, кратные ребра не допускаются;
- Ребрам присвоены веса из промежутка  $[0; 1]$ .

Также зафиксируем некоторую T-норму.

Для произвольной функции  $k$ -значной логики  $f(x_1, \dots, x_n)$  значение функции принадлежности  $\mu(f)$  определяется следующим образом: из множества вершин графа выделяется подмножество  $\{(\alpha, f(\alpha)), \alpha \in E_k^n\}$ . Значением  $\mu(f)$  является значение выбранной T-нормы от весов всех ребер подграфа, индуцированного этим подмножеством вершин.

Приведем некоторые наиболее распространенные типы нечетких условий, в том числе и описанные в работе [5], и способ их реализации в нашем подходе.

*Условие на значение в точке* 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).

Условие задается точкой  $\alpha$  и функцией принадлежности  $\mu : E_k \rightarrow [0; 1]$ . Для произвольной функции  $k$ -значной логики  $f(x_1, \dots, x_n)$  степень выполнения такого условия равна  $\mu(f(\alpha))$ . Условие может быть строгим – функция  $\mu$  принимает значение 1 в единственной точке и равна нулю в остальных, или нестрогим ( $\mu$  принимает положительные значения в нескольких точках).

Например, условие «на нулевом наборе функция должна принимать нулевое значение» будет задаваться набором  $\alpha = (0, \dots, 0)$  и функцией  $\mu(x)$ , принимающей значение 1 при  $x = 0$  и значение 0 в остальных случаях. Условие «на нулевом наборе функция должна принимать малое значение» может задаваться набором  $\alpha = (0, \dots, 0)$  и функцией  $\mu(x)$  следующего вида:

$x$	0	1	2	3	...	$k-1$
$\mu(x)$	1	0.8	0.4	0	...	0

В нашей графовой модели условие задается петлями на вершинах  $\{(\alpha, a), a \in E_k\}$ , с весами  $\mu(a)$  (если  $\mu(a) = 1$ , петля не добавляется).

*Локальное условие по одной переменной* 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.

Условие описывается матрицей  $\mu_{ij}, i, j \in E_k$ . Элемент матрицы  $\mu_{ij}$  определяет степень, с которой пара значений функции  $f(a)$  и  $f(a+1)$  удовлетворяет условию, если  $f(a) = i$  и  $f(a+1) = j$ . Например, условие «функция  $f(x)$  строго монотонна» задается следующей матрицей:

$f(i+1)$ $f(i)$	0	1	2	3	4	5
0	0	1	1	1	1	1
1	0	0	1	1	1	1
2	0	0	0	1	1	1
3	0	0	0	0	1	1
4	0	0	0	0	0	1
5	0	0	0	0	0	0

а условие «При увеличении  $x$  функция  $f$  слабо возрастает» может быть задано следующей матрицей ( $k = 6$ ):

$f(i+1)$ $f(i)$	0	1	2	3	4	5
0	0.2	1	1	0.6	0.2	0
1	0	0.2	1	1	0.6	0.2
2	0	0	0.2	1	1	0.6
3	0	0	0	0.2	1	1
4	0	0	0	0	0.2	1
5	0	0	0	0	0	0.2

Степень выполнения условия для функции будет вычисляться как некоторая T-норма от степеней выполнимости условия на всех наборах, отличающихся на единицу по соответствующей переменной. Например, если  $T(x, y) = xy$ , то для функции от одной переменной  $f(x)$

$x$	0	1	2	3	...	$k-1$
$f(x)$	1	0.8	0.4	0	...	0



степень выполнимости условия «слабого возрастания», заданного выше, будет равна:  $\mu(f) = \mu_{f(0)f(1)} \cdot \mu_{f(1)f(2)} \cdot \mu_{f(2)f(3)} \cdot \mu_{f(3)f(4)} \cdot \mu_{f(4)f(5)} = \mu_{01} \cdot \mu_{12} \cdot \mu_{22} \cdot \mu_{24} \cdot \mu_{45} = 1 \cdot 1 \cdot 0.2 \cdot 1 \cdot 1 = 0.2$ .

В графовой модели условие задается следующим образом: для каждой пары наборов  $(\alpha, \beta) : \alpha, \beta \in E_k^n, \alpha_i = \beta_i, i \neq j, \beta_j = \alpha_j + 1$ , где  $j$  — номер переменной, по которой строится условие, и каждой пары  $(a, b) : a, b \in E_k$  проводится ребро, соединяющее вершины  $(\alpha, a)$  и  $(\beta, b)$  и имеющее вес  $\mu_{a,b}$ .

Локальное условие по нескольким переменным определяет желаемое поведение функции на наборах, отличающихся на единицу по нескольким выбранным переменным. Условия такого типа могут задаваться высказываниями: «при совместном возрастании второй и четвертой переменных функция слабо возрастает» и т.п. Условие задается аналогично предыдущему с той лишь разницей, что ребра проводятся между вершинами с несколькими отличающимися на единицу координатами.

Мы можем комбинировать условия с помощью логических операций «И» и «ИЛИ». Итак, пусть задано множество вершин  $\{(\alpha, f(\alpha)), \alpha \in E_k^n, a \in E_k\}$  и два множества взвешенных ребер, соответствующих нечетким условиям. Конъюнкции двух условий будет соответствовать объединение множеств ребер. В случае если ребро входило в оба исходных множества, его весом в конъюнкции будет некоторая Т-норма исходных весов. Дизъюнкции двух условий будет соответствовать пересечение множеств ребер. Веса ребер в дизъюнкции будут вычисляться как некоторая Т-норма весов соответствующих ребер условий-аргументов.

Теперь мы формализуем задачу выбора адекватного оператора агрегирования и докажем ее труднорешаемость в общем случае. Итак, нечеткое условие на функцию  $k$ -значной логики  $f(x_1, \dots, x_n)$  ( $k$  фиксировано,  $n$  может быть произвольно большим) задано графом  $G$  с множеством вершин  $V_G = \{(\alpha, f(\alpha)), \alpha \in E_k^n, a \in E_k\}$ , некоторым множеством ребер  $E_G$  и некоторыми весами  $\mu_e \in [0; 1], e \in E_G$ . Нам нужно выбрать в этом графе подмножество вершин  $V'$  со следующими свойствами:

1.  $\forall \alpha \in E_k^n \exists! a \in E_k : (\alpha, a) \in V'$ ;
2. Определим функционал на графах с взвешенными ребрами:

$$\mu(G) = \bigwedge_{e \in E_G} \mu(e),$$

где  $\bigwedge$  — Т-норма. Тогда из всех графов, индуцированных подмножествами вершин, удовлетворяющих свойству 1, граф  $G'$ , индуцированный подмножеством  $V'$ , максимизирует этот функционал.

Перед доказательством труднорешаемости этой задачи, дополнительное ограничение. Разрешим весам принимать только значение 0. Во-первых, как будет показано ниже, даже при таком «предельном» случае задача остается NP-полной. Во-вторых, так как все Т-нормы принимают одинаковые значения, если аргументы принимают только значения 0 или 1, то мы таким образом сможем доказать труднорешаемость задачи для всех Т-норм сразу. В-третьих, как будет показано ниже, к задаче в такой постановке могут быть сведены некоторые смежные задачи. Среди них: является ли нечеткое условие внутренне противоречивым (т.е.

таким, что ему не удовлетворяет ни одна функция), является ли нечеткое условие строго выполнимым (т.е. существует ли функция, удовлетворяющая ему со степенью 1), а также вопрос о существовании функции, удовлетворяющей условию с произвольной заданной степенью в случае, когда в качестве Т-нормы выступает минимум.

**Теорема 2.** Следующая задача NP-полна: для данного графа с множеством вершин  $V_G = \{(\alpha, f(\alpha)), \alpha \in E_k^n, a \in E_k\}$  и некоторым множеством ребер требуется проверить, существует ли такое подмножество вершин, удовлетворяющее свойству 1, что индуцированный им подграф не имеет ни одного ребра.

*Доказательство.* Принадлежность задачи классу NP очевидна. Для доказательства NP-трудности сведем к нашей задаче классическую NP-полную задачу РАСКРАШИВАЕМОСТЬ ГРАФА [12] – для данного графа и натурального  $k$  проверить, существует ли у графа правильная вершинная раскраска в  $k$  цветов. Эта задача эквивалентна задаче о выполнимости следующей «конъюнктивной» формы с  $k$ -значными аргументами:  $\bigwedge_{\alpha\beta \in E} (x_\alpha \neq x_\beta)$  (значение переменной  $x_\alpha$  соответствует цвету вершины  $\alpha$ , множитель  $(x_\alpha \neq x_\beta)$  соответствует ребру  $\alpha\beta$ ). С помощью тождества  $(x_\alpha \neq x_\beta) = \bigwedge_{i \in E_k} ((x_\alpha \neq i) \vee (x_\beta \neq i))$  сведем задачу о выполнимости этой формы к задаче о выполнимости формы  $\bigwedge_{\alpha\beta \in E} ((x_\alpha \neq a_\alpha) \vee (x_\beta \neq x_\beta))$ , которая очевидным образом сводится к нашей задаче (значение переменной  $x_\alpha$  соответствует выбору вершины  $(\alpha, x_\alpha)$ , а множитель  $((x_\alpha \neq a_\alpha) \vee (x_\beta \neq x_\beta))$  соответствует ребру между вершинами  $(\alpha, a)$  и  $(\beta, b)$ ). Теорема доказана.

**Следствие 1.** NP-полны задачи проверки противоречивости и строгой выполнимости нечеткого условия.

**Следствие 2.** Для любой Т-нормы задача проверки существования функции, удовлетворяющей данному нечеткому условию со степенью не меньше заданной, NP-полна.

**Следствие 3.** При справедливости гипотезы  $P \neq NP$  не существует приближенного полиномиального алгоритма, решающего задачу выбора оператора агрегирования с результатом, отклоняющимся от оптимального не более чем в фиксированное число раз.

#### 4. 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, приведенные в разделе 2.2, а также демонстрируются различные подходы, диктуемые особенностями задач (масштабом контролируемой ситуации, особенностями используемых данных и т.д.).

##### 4.1 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 недекларированной ядерной активности стран-участниц. Для выявления недекларированной ядерной деятельности стран МАГАТЭ имеет политический мандат стран-участниц и механизмы реализации контрольных функций. 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 несоответствия 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.

Если в начале своей деятельности Агентство основное внимание уделяло учету и контролю материальных потоков, связанных с ядерными материалами, то, начиная с 1991 года, по решению Совета Безопасности ООН, Совет управляющих МАГАТЭ разработал и одобрил множество дополнительных мер, обеспечивающих доступ сотрудников Агентства к информации и объектам стран-участниц. В рамках такой усиленной системы гарантий (strengthened safeguards) предполагается, что для создания ядерного оружия кроме наличия ядерного материала страна должна обладать специальными технологиями, научной и экспериментальной базой, кадрами. Эксперименты могут оставлять следы во внешней среде; научные и технологические интересы страны могут быть выявлены на основе анализа публикаций и участия ученых и специалистов в международных конференциях; непропорциональное развитие определенных областей экономики страны также может быть индикатором наличия теневой ядерной программы.

Ясно, что для такого 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:

1. Является ли официальная ядерная программа страны взвешенной и последовательной?
2. Находятся ли реальные действия страны в ядерной области, особенно типы и объемы ядерных материалов и места их хранения и переработки, в соответствии с декларированными (официальными)? Подтверждается ли это анализом образцов окружающей среды (почва, вода, воздух)?

3. Находятся ли в соответствии производство, импорт, накопленные запасы ядерных материалов и объявленная в декларации официальная ядерная программа страны?
4. Совместим ли реальный импорт технологий, оборудования, специфических неядерных материалов и объявленная в декларации официальная ядерная программа страны?
5. Является ли существующий статус закрытых или выведенных их эксплуатации предприятий подтверждением официальной политики страны в ядерной области?
6. Находится ли в соответствии существующий уровень и научных и технологических разработок в связанных с ядерной областью областях науки и технологий в соответствии с планами развития ядерной области, объявленной в декларации 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», 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 нескольких 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.

Физическая модель имеет иерархическую структуру. Выделяются уровень технологий, уровень групп процессов, уровень процессов и уровень оборудования. Каждый элемент модели описан в текстовом виде (вход, выход, преобразование) и затем характеризуется в терминах индикаторов существования. Индикаторы

могут относиться к оборудованию, ядерным и неядерным материалам, состоянию окружающей среды, требованиям для определенных технических навыков, и т.д. Всего выделено 935 таких индикаторов, «привязанных» к элементам модели с указанием «силы» индикатора в терминах «сильный», «средний» и «слабый».

#### *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 надежным 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 DISNA 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 контролировать 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 хранение в базе данных всех документов (ссылок на документы), оцененных экспертом, с указанием их связи с элементами модели ядерной деятельности страны;
- обеспечить эксперта МАГАТЭ инструментом для ретроспективного анализа изменений оценок каждого элемента модели с возможностью просмотра соответствующего документа или ссылок на него;
- 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 [16]

#### *4.2 Прогнозирование риска развития сердечно-сосудистых заболеваний*

Известно, что особенное внимание в России и в мире уделяется проблемам, связанным с атеросклеротическими заболеваниями, при которых холестерин, продукты жизнедеятельности клеток, жировые субстанции, кальций, фибрин «налипают» на стенки артерий, в результате чего артерии утолщаются и отвердевают. Таким образом, ограничивается поток крови в артериях, нарушается гемодинамика, что в свою очередь, может вызывать повышение артериального давления, возникновение тромбов и т.д. Проблема атеросклеротических заболеваний остро стоит и в западных странах, где на здравоохранение выделяются существенные средства, для их лечения прикладываются огромные усилия государственных и частных структур, фармацевтические компании вкладывают в создание эффективных лекарств миллиарды долларов. И, по-видимому, не случайно: по данным American Heart Association, в США от сердечно-сосудистых заболеваний каждые 33 секунды умирает один человек. Россия – страна с одним из самых высоких показателей смертности от сердечно-сосудистых заболеваний. При этом финансирование, которое получают ведущие медицинские центры страны, занимающиеся проблемой атеросклероза и его последствий, явно недостаточно для борьбы с этим заболеванием. Поэтому перед институтом профилактической медицины МЗ РФ встала естественная задача о разработке методик и об отслеживании на новом технологическом уровне факторов риска атеросклеротических заболеваний



у населения России, прогнозирование дальнейшей динамики заболеваний этого класса в различных его возрастных группах и выработка наиболее экономически эффективных способов уменьшения риска.

Эта задача хорошо вписывается в класс задач, решаемых на основе технологий информационного мониторинга: имеются специалисты высокого уровня, обладающие достаточными экспертными знаниями по проблемам атеросклеротических заболеваний, а также накоплена обширная база, достаточная для построения содержательной модели предметной области. Наличие такого рода факторов и позволило подойти к решению проблемы отслеживания факторов риска атеросклеротических заболеваний именно при помощи технологии информационного мониторинга.

#### *Построение структурной модели*

На первом этапе требуется построить структурную модель уровня риска сердечных заболеваний. Для этого, в результате серии встреч со специалистами Института профилактической медицины, были выделены и описаны основные факторы риска, определены шкалы их значений, заданы их характеристики.

В настоящее время насчитывают более 30 факторов, действие которых влияет на риск возникновения и развития атеросклероза и его осложнений. С учетом критериев полноты и доступности информации, после консультаций со специалистами Института профилактической медицины были выделены следующие ключевые факторы риска:

1. *Уровень холестерина в крови.* Холестерин транспортируется в крови внутри липопротеидов. Существуют липопротеиды высокой плотности (ЛВП – так называемый «хороший» холестерин) и низкой плотности (ЛНП – «плохой» холестерин). Считается, что именно ЛНП оказывают большое влияние на риск заболевания атеросклерозом.
2. *Уровень триглицеридов (ТГ).* Триглицериды сами по себе не играют решающей роли в развитии сердечных заболеваний. Однако они имеют большое значение в совокупности с другими факторами риска. На практике высокие уровни триглицеридов очень часто сочетаются с низкими уровнями холестерина высокой плотности, то есть «хорошим холестерином», и это указывает на то, что частички жира поступают в кровь, что является причиной болезни сосудов. Поэтому высокий уровень триглицеридов должен восприниматься как сигнал опасности.
3. *Артериальное давление.* Возникновение и темпы развития атеросклероза тесно связаны с повышенным уровнем артериального давления крови (гипертоническая болезнь). Артериальное давление делится на два типа: систолическое (верхнее) и диастолическое (нижнее).
4. *Курение.* Может значительно повысить риск атеросклероза.
5. *Диабет.* Диабет также является важным негативным фактором риска.
6. *Физическая активность.* Значительно снижает риск сердечных заболеваний.

7. *Наследственность.* «Плохая» наследственность способствует развитию сердечных заболеваний и их осложнений.
8. *Питание.* Оказывает существенное влияние на риск развития сердечных заболеваний. Правильное питание может практически исключить такой риск, тогда как неправильное, особенно при неблагоприятной комбинации других ключевых факторов, значительно его увеличить. Фактически, питание вызывает изменение уровня холестерина и триглицеридов в крови, а также артериального давления, т.е. влияет на риск опосредованно, через другие факторы. Этот фактор представляет собой еще один пример ситуации, когда лингвистическое описание является значительно более удобным. Правильность питания – это агрегированный показатель, учитывающий различные виды питания.
9. *Body mass index (BMI).*  $BMI = \frac{mass, kg}{(height, cm)^2}$  – чем выше этот индекс, тем выше риск.

Многие из описанных ключевых факторов риска взаимосвязаны между собой. Например, физическая активность существенно влияет на уровень холестерина в крови, на артериальное давление, а также на вес и, следовательно, на индекс массы тела. Подобные взаимосвязи, естественно, должны быть учтены и в модели. Таким образом, выстраивается иерархическая древовидная модель предметной области.

По мнению специалистов, получившаяся структурная модель вполне адекватно описывает предметную область. То есть она, во-первых, достаточно точная, а во-вторых, можно довольно быстро и дешево получить всю информацию по используемым факторам риска.

#### *Построение модели правил*

По завершении этапа построения структурной модели предметной области, необходимо перейти к построению поведенческой модели, то есть задать зависимости между различными факторами риска. Эти зависимости в нашем случае были сформулированы в виде лингвистических правил. Как уже обсуждалось ранее, строгие формулы для описания зависимостей между факторами риска не очень эффективны, поскольку точные цифры не представляют особой ценности, а служат только лишь как показатель уровня, порядка риска, при этом получение строгих зависимостей процедура трудоемкая и требует привлечения дополнительных специалистов по статистическим методам. Поэтому для описания зависимостей между факторами риска было решено воспользоваться именно лингвистическими правилами. Одним из их несомненных достоинств является также и то, что такие правила сможет сформулировать специалист в области атеросклеротических заболеваний, т.е. отпадает необходимость в привлечении специалистов в области статистических методов. Кроме того, при помощи лингвистических правил можно достаточно адекватно описать зависимости между выбранными факторами риска.

#### *Инициализация*

После того, как была построена модель предметной области (структурная и поведенческая), следует приступить к сбору необходимой информации. Для этого

Институт профилактической медицины планирует провести крупномасштабное анкетирование населения России. В ходе анкетирования у респондентов будут взяты соответствующие анализы на уровень холестерина в крови, измерено артериальное давление, выяснено качество их питания, вредные привычки (курение, алкоголь) и т.д., то есть будут оценены все факторы риска, включенные в модель.

Наконец, когда вся информация собрана, ее необходимо ввести в модель. Это осуществляется следующим образом: сначала выделяется определенная группа населения, объединенная по некоторому признаку (например, по полу или возрасту), которая в данный момент интересует специалистов. Затем данные анкетирования о значениях факторов риска по выбранной группе усредняются и вносятся в качестве факторов риска в построенную модель.

### *Результаты*

Функционал системы аналогичен функционалу системы DISNA (см. раздел 1.4). Более подробное описание этого проекта приведено в работе [1]

#### *4.3 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 [13,14]. 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 улучшить возможности коллектива разработчиков оптимальным образом при заданном ресурсном обеспечении?

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

Для повышения надежности, 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 приведенные 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. Дело в том, что компания Cadence Design Systems ([www.cadence.com](http://www.cadence.com)) – лидер рынка САПР (систем автоматизированного проектирования) микроэлектроники – имеет необходимый опыт (модель «виртуальной САПР» Cadence VCAD) и потребность в такой системе как инструменте совершенствования услуг и сервисов VCAD. Описываемый ниже прототип разработан с привлечением экспертов компании и находится в опытной эксплуатации в компании.

### *Cadence VCAD*

Кратко, идея Cadence VCAD [15] заключается в следующем. Так как на сегодняшний день процесс разработки электронного устройства чрезвычайно сложен и во многом слабо формализован, многие, даже крупные компании-разработчики могут не обладать достаточными ресурсами (опытными руководителями, компетентными в конкретной области разработчиками, программным и аппаратным обеспечением) для успешного выполнения проекта. В этом случае, недостающие ресурсы (как людские, так и аппаратные и программные) обеспечивает служба VCAD, «виртуально расширяя» отдел разработки компании. В связи с этим естественно возникает вопрос оптимального с точки зрения качества конечного результата и объема затрат привлечения и распределения ресурсов. Именно этот круг вопросов (оценка имеющихся ресурсов, привлечение дополнительных, мониторинг выполнения проекта) и решает описываемая система.

### *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 (например, для навыков – знакомство с конкретными программами, опыт участия в похожих проектах и т.п.). Модель содержит более 50 элементов и более 70 связей, объединенных в 5 уровней иерархии.

### *Создание поведенческой модели*

В этом проекте ввод лингвистических правил осуществлялся вручную – эксперт не испытывал затруднений в указании значения для любой комбинации



Рис. 5: Структура модели

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 (например: «если навык разработчика *высокий* и программное обеспечение *хорошее*, то качество результата *высокое*»)
- В большинстве случаев набор правил обладал свойством монотонности: большие значения факторов-параметров влекли за собой большее 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 остается ввести их в систему.

### Результаты

На момент написания этой работы функциональный прототип системы успешно прошел первичное тестирование и перешел в стадию опытной эксплуатации. Более подробное описание этого проекта приведено в работе [3].

#### 4.4 Возможные области применения технологии информационного мониторинга

Приведенные 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.

## 5. 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, здравоохранение и менеджмент.

### Список литературы

- [1] Ахмеджанов Н.М., Жукоцкий А.В., Кудрявцев В.Б., Оганов Р.Г., Расторгуев В.В., Рыжов А.П., Строгалов А.С. Информационный мониторинг в задаче прогнозирования риска развития сердечно-сосудистых заболеваний. Интеллектуальные системы, Том 7, выпуск 1-4, 2003, с. 5-38.
- [2] Заде Л.А. Понятие лингвистической переменной и его применение к принятию приближенных решений. М.: Мир, 1976. 165с.
- [3] Лебедев А.А., Рыжов А.П. Оценка и мониторинг проектов разработки высокотехнологичных изделий на примере микроэлектроники. Интеллектуальные системы, Том 11, выпуск 1-4, 2008 г., с. 55-82.
- [4] Кудрявцев В.Б. Функциональные системы. Издательство Московского Университета, 1982. 158с.
- [5] Рыжов А.П. Об агрегировании информации в нечетких иерархических системах. Интеллектуальные системы, Том 6, выпуск 1-4, 2001, с. 341-364.



- [6] Рыжов А.П. О степени нечеткости размытых характеристик. Математическая кибернетика и ее приложения в биологии. Под редакцией Л.В.Крупинского, С.В.Яблонского, О.Б.Лупанова. М., Издательство МГУ, 1987, с. 60-77.
- [7] Рыжов А.П. Оценка степени нечеткости и ее применение в системах искусственного интеллекта. Интеллектуальные системы. Т.1, Вып.1-4, Москва, МНЦ КИТ, 1996, с. 95-102.
- [8] А.П. Рыжов, В.В. Расторгуев. 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. Том 1, Москва, Физматлит, 2006 с. 70-81.
- [9] Яблонский С.В. Основные понятия кибернетики. Проблемы кибернетики. 1959. Вып.2, с.7-38.
- [10] Even S., Itai A., Shamir A. On the complexity of timetable and multicommodity flow problems. SIAM J: Comput. 5, 1976, pp.691-703.
- [11] Kaplan R.S., Norton D.P. The balanced scorecard: measures that drive performance. Harvard Business Review Jan - Feb 1992 pp. 71-80.
- [12] Karp R.M. Reducibility among combinatorial problems. Complexity of computer computations, 1974, Plenum Press, NY, pp 85-103.
- [13] Libecap Garry D. University entrepreneurship and technology transfer: process, design, and intellectual property. ELSEVIER, 2005. 311p.
- [14] Martin Michael J.C. Managing innovation and entrepreneurship in technology based firms. NY: John Wiley & Sons, Inc., 1994. 402p.
- [15] 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).
- [16] 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. 122p.
- [17] Usama M. Fayyad (Ed.). Advances in Knowledge Discovery and Data Mining. MIT Press, 1996. 560p.
- [18] Zadeh L.A. Fuzzy sets. Information and Control, 1965, v.8, pp. 338-353.