

## СПОСОБЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ В СЛОЖНЫХ СИСТЕМАХ

Ю.А. КРАВЧЕНКО, А.А. ЛЕЖЕБОВ, Д.Ю. ЗАПОРОЖЕЦ

ФГАОУ ВПО Южный федеральный университет  
Технологический институт в г. Таганроге  
347928, г. Таганрог, пер. Некрасовский, 44, ГСП-17А  
E-mail: [rector@tti.sfedu.ru](mailto:rector@tti.sfedu.ru)

*В статье приведена информация об использовании MDX-запросов при интеллектуальном анализе данных для задач принятия решений и об использовании аппарата имитационного моделирования сетей Петри при интеллектуальном анализе данных для задач принятия решений в сложных динамических системах. Рассматриваются преимущества использования имитационных моделей при анализе априорной информации о состоянии сложного объекта и динамики его развития в условиях воздействия факторов внешней среды. Рассмотрен абстрактный пример функционирования сети Петри на трех множествах априорной информации. Описаны перспективы применения данного подхода для представления и извлечения знаний из различных предметных областей.*

**Ключевые слова:** принятие решений, неопределенность, извлечение знаний, имитационное моделирование, интеллектуальный анализ данных.

### ВВЕДЕНИЕ

Объектом данной работы являются анализ технологии OLAP, построение многомерных структур обработки данных и запросов к ним на языке MDX. OLAP (online analytical processing) – технология обработки информации, включающая составление и динамическую публикацию отчётов и документов для аналитиков. MDX (multidimensional expressions) – язык многомерных запросов для работы с кубами [1].

В случае экстраполяции свойств современных сложных объектов проверка условия достоверности оказывается настолько проблематичной, что вынуждает идти по пути использования математических моделей, которые могут быть воплощены в программном обеспечении испытательных комплексов. Применение математической модели позволяет широко использовать возможности ЭВМ в контуре контроля и управления при проверке сложного объекта. Суть контроля заключается в сравнении состояний сложного объекта и модели после отработки каждого управляющего воздействия в соответствии с программой испытаний.

Сложность объектов не позволяет строить для них абсолютно адекватные модели. Математическая модель, будучи формализованным абстрактным описанием объекта, в состоянии охватить только основные, характерные его свойства, оставляя в стороне несущественные факторы.

### 1. Возможности языка MDX запросов

Актуальность OLAP-технологий обусловлена их практической значимостью для анализа больших объемов данных. Основные особенности OLAP:

- работает с многомерными структурами;
- получает информацию об аномалиях, закономерностях и трендах в базе данных без выяснения смыслового значения данных;

- обеспечивает быстрый доступ к результатам;
- использует заранее рассчитанные сводные данные;
- позволяет использовать различную агрегацию данных на каждом из уровней разбора запроса MDX;
- дает возможность отображать N-мерные срезы;
- позволяет получить необходимую нам информацию, не вдаваясь в подробности способа выдачи этой информации из базы данных [2].

В ходе исследования OLAP-технологии выделены основные этапы аналитического процесса.

1. Определение списка возможных запросов аналитика. Данная информация может быть получена непосредственно от аналитика либо выяснена путем общего анализа базы данных на предмет фактов и их зависимостей.

2. Анализ структуры базы данных и выборка необходимого и достаточного набора измерений и мер. Данный этап производится по информации из списка запросов аналитика. На основании этой информации строится представление базы данных, в которое включены только те таблицы и именованные запросы, которые содержат информацию, необходимую аналитику для построения отчетов. Сложность его построения напрямую зависит от степени денормализации исходной базы. Использование представления позволяет повысить безопасность системы, поскольку не происходит прямое обращение к исходной базе.

3. Создание многомерной структуры на основе полученного набора. Данный этап производится с использованием специального инструментария (в данном случае использовалась Business Intelligence Studio 2005). OLAP куб содержит информацию о способе группировки данных представления.

4. Выполнение запросов, построение отчетов, сбор статистической информации. Данный этап является финальным в аналитическом процессе. Для получения информации используется язык многомерных запросов MDX с возможностью различного агрегирования данных по различным созданным измерениям.

## 2. РАЗРАБОТКА ИМИТАЦИОННОЙ МОДЕЛИ АНАЛИЗА ДАННЫХ

В работе рассматривается взаимодействие внутренних моделей компонентов интеллектуальных информационных систем с помощью аппарата сетей Петри. Модели компонентов создаются на основе комбинации свойств интеллектуальных агентов. Каждый агент имеет механизм обработки входящих и исходящих сообщений, получения данных из входящих сообщений и преобразования исходящей информации в единый заданный формат.

Моделирование в сетях Петри осуществляется на событийном уровне. Определяются, какие действия происходят в подсистеме, какие состояния предшествовали этим действиям и какие состояния примет подсистема после выполнения действия. Выполнение событийной модели в сетях Петри описывает поведение подсистемы. Анализируя результаты выполнения, можно сказать о том, в каких состояниях пребывала или не пребывала подсистема, какие состояния в принципе не достижимы.

## 3. МОДЕЛИРОВАНИЕ НА ОСНОВЕ СЕТЕЙ ПЕТРИ

Сеть Петри определяется как четверка  $\langle P, T, I, O \rangle$ , где  $P$  и  $T$  – конечные множества позиций и переходов,  $I$  и  $O$  – множества входных и выходных функций. Другими словами, сеть Петри представляет собой двудольный ориентированный граф, в котором позициям  $P$  соответствуют вершины, изображаемые кружками, а переходам  $T$  – вершины, изображаемые утолщенными черточками; функциям  $I$  соответствуют дуги, направленные от позиций к переходам, а функциям  $O$  – дуги, направленные от переходов к позициям [3].

В сетях Петри вводятся объекты двух типов: динамические, которые изображаются метками (маркерами) внутри позиций, и статические, которые соответствуют вершинам сети Петри.

Маркировка – распределение маркеров по позициям. Маркеры могут перемещаться в сети. Каждое изменение маркировки называют событием, причем каждое событие связано с определенным переходом. События происходят мгновенно и одновременно при выполнении некоторых условий.

Каждому условию в сети Петри соответствует определенная позиция. Совершению события соответствует срабатывание перехода, при котором маркеры из входных позиций этого перехода перемещаются в выходные позиции. Последовательность событий образует моделируемый процесс [3].

В рассматриваемом случае начальными условиями для имитационной модели функционирования сложной системы будут служить три множества (рис. 1):

- 1)  $C$  – множество воздействий внешней среды;
- 2)  $A$  – множество внутренних характеристик системы;
- 3)  $A_1$  – множество управляющих воздействий на систему.

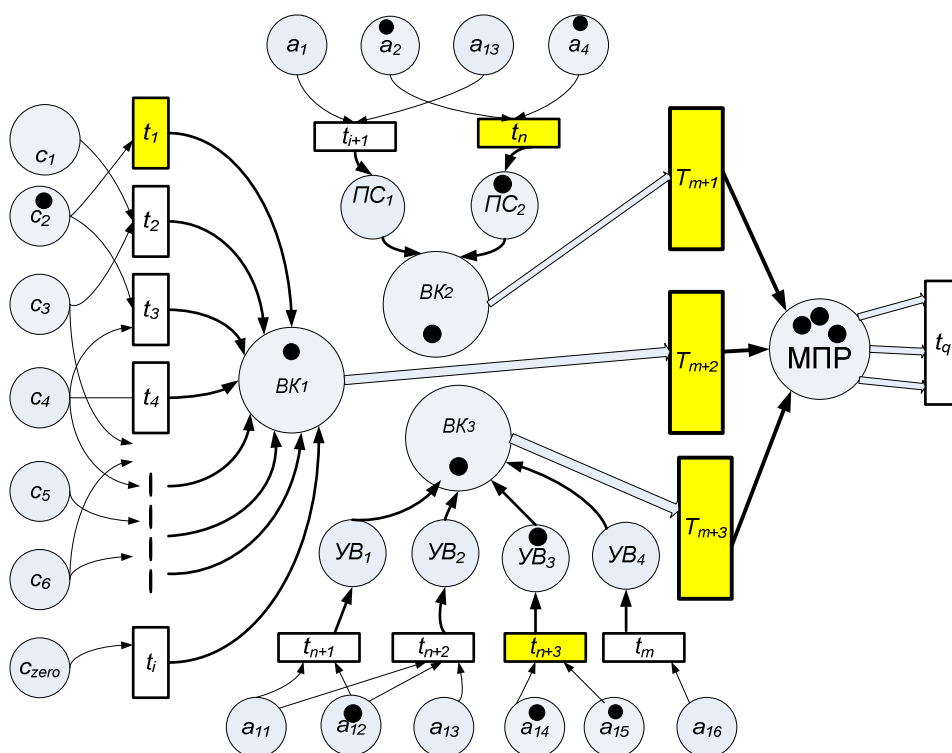
Предложим возможный вариант состава элементов данных множеств, мощности множеств выберем случайным образом, реализуя тем самым абстрактный пример функционирования сети. Множество  $C = \{c_1, c_2, c_3, c_4, c_5, c_6, c_{zero}\}$ , где  $c_1-c_6$  – информация о возможных воздействиях внешней среды,  $c_{zero}$  – вершина, необходимая теоретически для моделирования ситуации полного отсутствия выделенных воздействий, метка в ней появится только в том случае, если будут отсутствовать метки во всех остальных вершинах множества  $C$ .

Для учета всех возможных комбинаций наличия составляющих воздействия внешней среды в данном абстрактном примере необходимо создать  $2^6$  переходов. Причем дисциплина обслуживания переходов должна быть с абсолютным приоритетом в зависимости от количества входящих в переход дуг. С увеличением количества входящих в переход дуг его приоритет должен возрастать. Это необходимо для разрешения конфликтных ситуаций в пользу событий, фиксирующих наличие большего числа выделенных составляющих внешних воздействий.

Таким образом, открывшись, переход  $t_i$ ,  $i=\overline{1,2^6}$  отправит информацию о наличии составляющих воздействий внешней среды в вершину классификатора  $BK_1$ , наличие метки в которой позволит открыть переход  $t_{m+2}$  и передать полученные сведения в модуль принятия решения МПР (рис. 1).

С другой стороны, на принятие решений по управлению системой будет влиять комбинация меток в вершинах множества  $A = \{a_1, a_2, a_3, a_4\}$ , описывающих множество внутренних параметров системы. Причем в этом множестве необходимо учитывать как наличие отдельных характеристик, так и их комбинации, т.к. могут существовать недопустимые соотношения параметров управляемого сложного объекта. Определение допустимых наборов параметров управляемой системы происходит посредством передачи маркеров через переходы  $t_n$ ,  $n=\overline{i+1, i+2}$  к вершинам  $ПС_1$  или  $ПС_2$ , моделирующих выборки допустимых параметров системы. В дальнейшем эта информация попадет в вершину классификатора  $BK_2$  и далее к модулю принятия решения.

Подобным образом комбинация меток в вершинах множества  $A_1 = \{a_{11}, a_{12}, a_{13}, a_{14}, a_{15}, a_{16}\}$  описывает допустимые наборы управляющих воздействий на систему посредством передачи маркеров через переходы  $t_m$ ,  $m=\overline{n+1, n+4}$  к вершинам  $УВ_1, УВ_2, УВ_3, УВ_4$ . В дальнейшем эта информация попадет в вершину классификатора  $BK_3$  и далее к модулю принятия решения МПР (рис. 1).



**Рис. 1.** Абстрактный пример работы имитационной модели с выполнением условия достижимости сети

#### 4. АНАЛИЗ АПРИОРНОЙ ИНФОРМАЦИИ НА ОСНОВЕ ИМИТАЦИОННОЙ МОДЕЛИ

Покажем работу имитационной модели с выполнением условия достижимости, в качестве априорной информации будем использовать начальные условия, представленные на рис. 1.

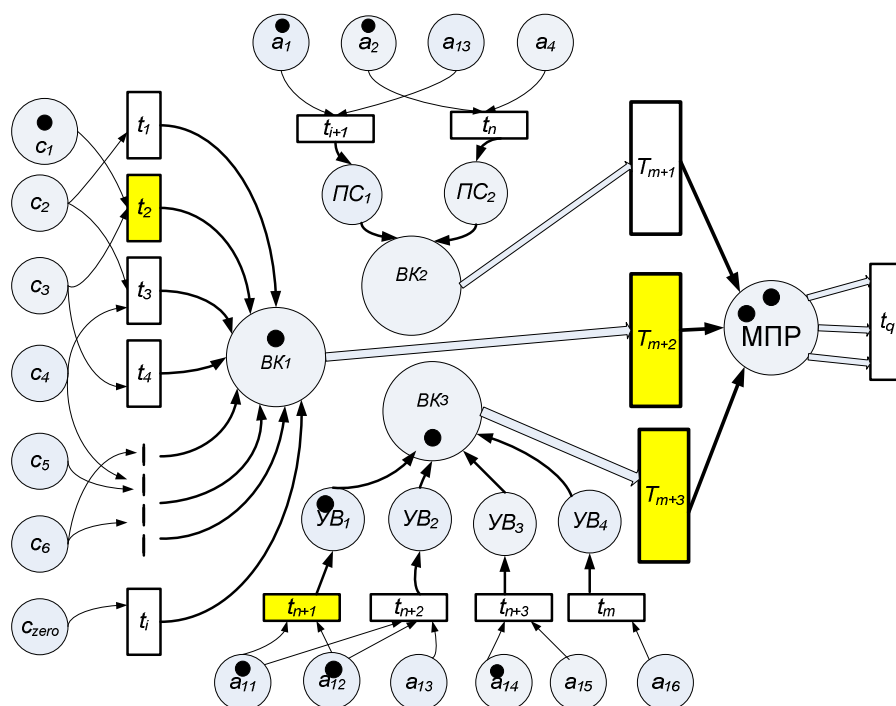
В данном случае начальные условия позволяют в ходе работы сети определить допустимое множество внутренних параметров системы (переход  $t_n$ , вершина  $PC_2$ ), допустимое множество управляющих воздействий (переход  $t_{n+3}$ , вершина  $UB_3$ ), параметры воздействия внешней среды (переход  $t_1$ , вершина  $BK_1$ ). Это позволяет через вершины  $BK_1, BK_2, BK_3$  открыть переходы  $T_{m+1}, T_{m+2}, T_{m+3}$ .

Результатом работы сети являются передвижение трех маркеров в вершину модуля принятия решения (МПР) и выработка им рекомендаций по возможным корректировкам управляющих воздействий. Дальнейшая возможность открытия виртуального перехода  $t_q$  позволяет сделать вывод о достижимости сети, что подтверждает решение задачи.

Покажем теперь работу модели на абстрактном примере, иллюстрирующем нарушение условия достижимости сети (рис. 2). В данном случае начальные условия позволяют в ходе работы сети определить допустимое множество управляющих воздействий (переход  $t_{n+1}$ , вершина  $UB_1$ ), параметры воздействия внешней среды (переход  $t_2$ , вершина  $BK_1$ ), но допустимое множество параметров системы не установлено, так как в вершину  $BK_2$  не переместился ни один маркер. Это произошло из-за того, что параметры системы, соответствующие вершинам  $a_1$  и  $a_2$ , не являются допустимой комбинацией.

Результатом работы сети является передвижение только двух маркеров в вершину модуля принятия решения. Поэтому виртуальный переход  $t_q$  не будет открыт, что позволяет сделать вывод о невозможности решения поставленной задачи. Корректировка управ-

ляющих воздействий на систему в данном случае будет проводиться экспертом на основе дополнительной оценки оперативной ситуации.



**Рис. 2.** Абстрактный пример работы имитационной модели с нарушением условия достижимости сети

### ЗАКЛЮЧЕНИЕ

Рассмотренный в статье подход имитационного моделирования принятия решений в интеллектуальных информационных системах использует интегрированный метод представления знаний. Это позволяет применять в исследованиях параметры из различных предметных областей для решения важных задач извлечения и обработки знаний. Подобная интеграция позволит оперативно отслеживать изменения в динамике развития системы и корректировать управляющие воздействия при изменении начальных условий.

Использование формальных методов (например, сетей Петри) для синтеза знаний с корректным содержанием может стать основой методов создания интеллектуальных информационных систем. Методы анализа воздействий внешней среды, множеств параметров системы и управляющих функций с использованием дерева достижимости и матричной теории сетей Петри позволят формализовать многие процессы, обладающие некоторой степенью неопределенности.

### ЛИТЕРАТУРА

1. Эндрю Дж. Браст, Стивен Форте. Разработка приложений на основе Microsoft SQL Server 2005. Мастер-класс / Пер. с англ. М.: Издательство «Русская Редакция», 2007.
2. Бергер А. Microsoft SQL Server Analysis Services, OLAP и многомерный анализ данных. СПб.: Издательство «БХВ-Петербург», 2007.
3. Норенков И.П., Кузьмик П.К. Информационная поддержка наукоемких изделий. CALS-технологии. М.: Изд-во МГТУ им. Н.Э. Баумана, 2002.
4. Nelson G. and J. Wright. Real Time Decision Support: Creating a Flexible Architecture for Real Time Analytics. DSSResources.COM, 11.18.2005.

## METHODS OF INTELLIGENT DATA ANALYSIS IN COMPLEX SYSTEMS

**Y.A. KRAVCHENKO, A.A. LEZHEBOKOV, D.Y. ZAPOROZHETS**

Southern Federal University, Technological Institute at Taganrog  
347928, Taganrog, 44, Nekrasovsky Lane, GSP-17A  
E-mail: [rector@tti.sfedu.ru](mailto:rector@tti.sfedu.ru)

This article contains information about application of MDX requests for the intellectual data analysis for support of decision-making and about using the simulation Petri nets at intellectual analysis for decision-making problems in complex dynamic systems. The advantages of simulation models in the analysis of a priori information about the state of a complex object and the dynamics of its development under the impact of environmental factors were described. We considered an abstract example of a network function on the three sets a priori information. We described the prospects of applying this approach to representation and extraction knowledge from different subject areas.

**Key words:** decision-making, uncertainty, knowledge extraction, simulation modeling, intelligent data analysis.

*Работа поступила 20. 05. 2012 г.*