

## **Вычислительный эксперимент**

*Предлагается один из возможных подходов к интеллектуализации процессов оптимизационно-имитационного моделирования, базирующийся на концепциях и технологиях, принятых в современной мировой и отечественной практике интеллектуального анализа данных. Рассматриваются особенности реализации такого подхода применительно к системе моделирования NEDISOPT\_D.*

© В.Б. Бигдан, 2009

УДК 681.3.06+519.8

В.Б. БИГДАН

## **ОСОБЕННОСТИ РЕАЛИЗАЦИИ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА РЕЗУЛЬТАТОВ ИМИТАЦИОННЫХ ЭКСПЕРИМЕНТОВ**

**Введение.** Одной из основных тенденций, которые наметились в последние годы в зарубежной практике имитационного моделирования, есть развитие и широкое распространение методологии Data Farming. Отличительной особенностью данной методологии является интеграция возможностей методов имитационного моделирования, различного рода оптимизационных стратегий, методов интеллектуального анализа данных и технологий распределенных вычислений, которые реализуются на высокопроизводительных платформах вычислительной техники, в частности, кластерных или сетевых архитектурах [1, 2]. При этом значительное повышение эффективности процессов имитационного моделирования достигается за счет использования широкого спектра методов интеллектуального анализа результатов оптимизационно-имитационных экспериментов. Традиционно в системах имитационного моделирования анализ накопленной в процессе моделирования статистики базируется на концепции "усреднения по выборке". Такого рода классический анализ поддерживается изобразительными средствами базовых языков моделирования и осуществляется регулярно по завершении отдельных прогонов соответствующей имитационной модели. Разнородность, многомерность и большой объем выходной информации, получаемой в процессе реализации оптимизационно-имитационных экспериментов согласно методологии Data Farming, не позволяют

проводить эффективный анализ последней на основе классических методов прикладной статистики, а требуют использования средств интеллектуального анализа данных (ИАД). Прежде всего, здесь идет речь об оценке зависимостей целевых показателей эффективности соответствующих систем или процессов (как правило, представляемых откликами моделей и значениями функции цели) от факторов, задающих характеристики оцениваемых альтернатив.

Интеллектуальный анализ данных является мультидисциплинарной областью информационных технологий, которая развивается на базе достижений прикладной статистики, распознавания образов, методов искусственного интеллекта, теории экспертных систем, теории баз данных, поиска и оперативной обработки данных, технологий визуализации данных и др. [3, 4].

Современные методы и средства ИАД базируются на концепциях технологий OLAP (on line analytical processing), Data Mining и визуализации данных. Хранилища данных являются важной составляющей систем ИАД, поскольку хранят объединенные и согласованные (консолидированные) данные.

Возможности технологии OLAP используются для предварительного анализа и оперативной обработки многомерных данных. Средства OLAP предназначены в основном для быстрого составления отчетов по консолидированным показателям процессов в различных разрезах и произвольной глубиной погружения в оперативные данные. Задачи систем поддержки технологии OLAP осуществляются в терминах соответствующей предметной области.

Применительно к процессам моделирования методы, алгоритмы и средства технологии Data Mining используются для поиска и обнаружения скрытых, ранее не известных закономерностей и знаний на основе анализа исторических данных, составляющих основу опыта моделирования в рассматриваемой области исследования.

Наиболее известные в Украине системы моделирования, в том числе системы распределенного моделирования, разработанные киевскими специалистами под руководством д-ра техн. наук В.В. Литвинова, чл.-кор. Т.П. Марьяновича, д-ра техн. наук В.Н. Томашевского, поддерживают основные концепции методологии Data Farming, но содержат слабо развитые средства интеллектуального анализа данных. Поэтому проблема расширения функциональных возможностей отечественных систем имитационного моделирования за счет методов интеллектуального анализа данных является актуальной.

**Постановка задачи и цель исследования.** Современные методы и программные средства поддержки интеллектуального анализа данных классифицируются следующим образом: по типам анализируемых данных; по типам решаемых задач; по методам анализа; по классам используемых алгоритмов; по областям приложений.

С помощью программного обеспечения ИАД решаются следующие задачи: классификация (отнесение объекта к наперед определенной категории); прогнозирование (определение значений неизвестных объектов на основе анализа атрибутов известных объектов); ассоциация (определение зависимостей между атрибутами); кластеризация (выделение компактных подгрупп подобных объектов); дискриминантный анализ (выявление атрибутов, которыми отличаются две или более групп объектов).

Наиболее распространенные методы анализа в системах ИАД: традиционные статистические методы (статистические пакеты SAS, SAS Institute; SPSS, SPSS; STATGRAPICS, Manugistics; STATISTICA, MATHEMATIC), нейронные сети (BrainMaker, CSS; NeuroShell, Ward Systems Group; OWL, HyperLogic); системы рассуждений на основе аналогичных случаев (метод ближайшего соседа – KATE tools, Acknosoft, Франция; Pattern Recognition Workbench, Unica, США), деревья решений (RuleQuest, Австралия; Clementine, Integral Solutions, Великобритания; SIPINA, University of Lyon, Франция; IDIS, Information Discovery, США; KnowledgeSeeker, ANGOSS, Канада), эволюционное программирование (PolyAnalyst, Megaputer Intelligence), генетические алгоритмы (GeneHunter, Ward Systems Group), алгоритмы ограниченного перебора (WizWhy, WizSoft), методы МГУА (методы группового управления аргументами – NeuroShell, Ward Systems Group), методы и средства графического отображения многомерных данных (визуализация данных – DataMiner 3D, Dimension5) [5].

Целью исследования является разработка методов и средств поддержки ИАД результатов оптимизационно-имитационных экспериментов, реализуемых в рамках системы NEDISOPT\_D, разработанной в Институте кибернетики имени В.М. Глушкова НАН Украины [6, 7]. Здесь важный этап – определение и разработка структурной организации хранилища данных, содержащего опыт моделирования в форматах специально разработанных шаблонов-моделей для объектов входной и выходной информации системы NEDISOPT\_D, а также соответствующий пользовательский интерфейс.

Предполагается реализация указанных средств ИАД в виде следующего набора инструментариев: система 1 должна обеспечивать доступ к данным в хранилище и оперативную выдачу отчетов; система 2 разрабатывается для поддержки OLAP – анализа оперативных данных из хранилища; система 3 предназначена для анализа информации в соответствии с основными концепциями технологии Data Mining, прежде всего должны быть оценены взаимодействие и влияние факторов на результаты поиска оптимальных решений.

**Особенности реализации методов ИАД в среде системы NEDISOPT\_D.** Взаимодействие основных компонент системы NEDISOPT\_D и систем поддержки интеллектуального анализа результатов поиска оптимальных решений представлено на рисунке. Здесь MSCN – главный сценарий системы NEDISOPT\_D, управляющий процессами реализации оптимизационно-имитационных экспериментов в формате сессий моделирования; сценарии SCN\_OSi включают модули поддержки основных оптимизационных стратегий системы NEDISOPT\_D; сценарии SCNA\_L – управляют прогоном имитационных моделей SMA\_L; сценарий SCN\_SE контролирует условия завершения сессии моделирования; сценарий SCN\_RFORM обеспечивает обработку и выдачу результатов поиска оптимальных решений.

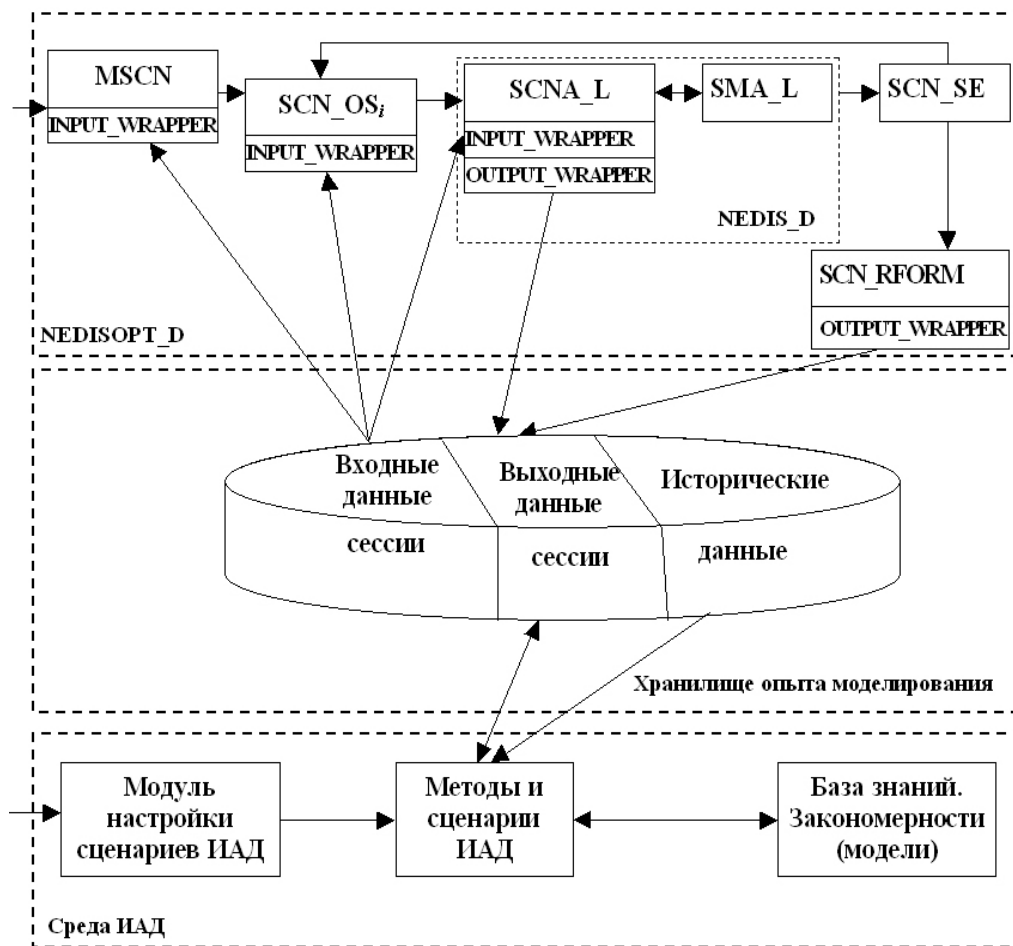


РИСУНОК. Схема поддержки методов ИАД для оценки результатов оптимизационно-имитационных экспериментов

Заметим, что исследования на основе системы NEDISOPT\_D проводятся в формате сессий моделирования. В хранилище данных (опыта моделирования) в формате шаблонов-моделей загружается входная информация сессий моделирования, выходные результаты текущих сессий моделирования и исторические данные. Процедуры поиска необходимых данных в хранилище опираются на библиотеку шаблонов.

Неотъемлемой компонентой всех типов сценариев являются специально разработанные wrapper-процедуры, которые обеспечивают перевод информации из форматов хранилища данных в форматы системы NEDISOPT\_D и, наоборот, в зависимости от направления потока данных.

Конфигурационные параметры сессии моделирования, являющиеся входными данными сценария MSCN, определяют конфигурацию сети и динамически формируемую цепочку сценариев, поддерживающих соответствующую схему поиска оптимальных решений.

Аналогичным образом модуль настройки сценариев ИАД на основе множества конфигурационных параметров для ИАД определяет динамически создаваемую цепочку сценариев, ориентированную на выполнение соответствующих задач интеллектуального анализа результатов моделирования.

Модули системы 1 обеспечивают поиск затребованной по заданию аналитика или разработчика модели информации с целью формирования соответствующих отчетов по сессиям моделирования.

Модули системы 2, опирающиеся на методы и средства OLAP-технологии, поддерживают решение таких задач, как верификация и валидация имитационных моделей, нахождение и устранение всякого рода неточностей и неопределенностей, касающихся объектов имитационной модели и вопросов определения корректных границ и уровней квантования для факторов, выбор размера начальной популяции для соответствующих оптимизационных стратегий. При этом обеспечивается удаление из хранилища шаблонов, не представляющих дальнейшего интереса для аналитика, эксперта или разработчика модели. Корректные с точки зрения пользователя системы NEDISOPT\_D результаты переносятся в раздел "исторические данные".

Модули системы 3 на основе методов и средств Data Mining осуществляют поиск скрытых, ранее неизвестных закономерностей на основе анализа исторических данных, содержащих накопленный опыт моделирования. Результаты поиска, определяемые как соответствующие модели или закономерности, сохраняются в базе знаний.

**Заключение.** Предложен подход к расширению функциональных возможностей системы распределенного моделирования NEDISOPT\_D средствами интеллектуального анализа, обеспечивающими анализ и обработку результатов направленного поиска оптимальных решений. К перспективным направлениям следует отнести расширение множества методов анализа данных и обеспечение возможностей композиции различных методов и подходов.

*В.Б. Бигдан*

#### ОСОБЛИВІСТІ РЕАЛІЗАЦІЇ МЕТОДІВ ІНТЕЛЛЕКТУАЛЬНОГО АНАЛІЗУ РЕЗУЛЬТАТІВ ІМІТАЦІЙНИХ ЕКСПЕРИМЕНТІВ

Пропонується один із можливих підходів до інтелектуалізації процесів оптимізаційно-імітаційного моделювання, який базується на концепціях та технологіях, що прийняті у сучасній світовій та вітчизняній практиці інтелектуального аналізу даних. Розглядаються особливості реалізації такого підходу стосовно системи моделювання NEDISOPT\_D.

*V.B. Bigdan*

PECULIARITIES OF REALIZATION OF DATA MINING METHODS  
FOR SIMULATION EXPERIMENTS RESULTS

One of the possible approaches to intellectualization of the processes of optimization-imitation simulation based on concepts and technologies adopted in the existing worldwide and domestic practice is proposed. Peculiarities of realization of such approach are considered in application to *NEDISOPT\_D* simulating system.

1. *Davis Dan M., Baer Garth D., Gottschalk Thomas D.* 1st Century Simulation: Exploiting High Performance Computing and Data Analysis // Interservice / Industry Training, Simulation, and Education Conference (IITSEC). – 2004. – Paper N 1517. – P. 1–14.
2. *Horne Gary E., Meyer Theodore E.* Data Farming: Discovering Surprise // Proceedings of the 2005 Winter Simulation Conference. – 2005. – P. 1082–1087.
3. *Шанот М.* Интеллектуальный анализ данных в системах поддержки принятия решений // <http://www.osp.ru/os/1998/01/179360/>
4. *Brady Thomas F., Yellig Edward.* Simulation data mining: a new form of computer simulation output // Proc. of the 2005 Winter Simulation Conf. – 2005. – Paper N 05-030. – P. 285–289.
5. *Дюк В.* Data Mining - интеллектуальный анализ данных. // [http://www.iteam.ru/publications/it/section\\_92/article\\_1448/](http://www.iteam.ru/publications/it/section_92/article_1448/) – 2003. – С. 1–13.
6. *Галаган Т.Н., Пенеляев В.А., Сахнюк М.А., Чёрный Ю.М., Шваб Н.Д.* О моделях сценариев распределенного поиска оптимальных решений // Компьютерная математика. – 2007. – № 2. – С. 144–156.
7. *Пенеляев В.А., Сахнюк М.А., Чёрный Ю.М., Шваб Н.Д.* К вопросу о реализации мета-эвристических стратегий оптимизации моделирования // Там же. – 2005. – № 2. – С. 26–33.

Получено 09.04.2009

**Об авторе:**

*Бигдан Вера Борисовна,*

научный сотрудник Института кибернетики имени В.М. Глушкова НАН Украины.