



II МЕЖВУЗОВСКАЯ НАУЧНО-ПРАКТИЧЕСКАЯ  
КОНФЕРЕНЦИЯ  
«БИЗНЕС-АНАЛИТИКА. ИСПОЛЬЗОВАНИЕ  
АНАЛИТИЧЕСКОЙ ПЛАТФОРМЫ DEDUCTOR  
В УЧЕБНОМ ПРОЦЕССЕ ВУЗА»



# Deductor

28 июня 2011 года  
г. Москва

**СБОРНИК МАТЕРИАЛОВ КОНФЕРЕНЦИИ**

Рязань 2011

**Бизнес-аналитика. Использование аналитической платформы Deductor в учебном процессе вуза: сборник материалов II Межвуз. науч.-практ. конф. – Рязань: ООО «Лаборатория баз данных», 2011. – 120 с.**

В сборник материалов II Межвузовской научно-практической конференции «Бизнес-аналитика. Использование аналитической платформы Deductor в учебном процессе вуза», проводимой BaseGroup Labs совместно с РГАУ-МСХА имени К.А. Тимирязева, включены работы преподавателей, активно использующих аналитическую платформу Deductor, из ряда вузов и организаций городов Москва, Ахтубинск, Нижний Новгород, Барнаул, Тула, Волжский, Омск, Чебоксары, Ровно, Омск, Уфа, Харьков.

Включенные в сборник статьи разделены на две секции: 1) опыт преподавания дисциплин с использованием аналитической платформы Deductor; 2) актуальные задачи бизнес-аналитики и их решение методами Data Mining.

**Организационно-программный комитет конференции**

Арустамов А.И., директор BaseGroup Labs;  
Ахметов Р.Г., д.э.н., профессор, декан экономического факультета РГАУ-МСХА им. К.А. Тимирязева;  
Паклин Н.Б., к.т.н., руководитель образовательных проектов BaseGroup Labs;  
Гатаулин А.М., д.э.н., чл. – корр. РАСХН, РГАУ-МСХА им. К.А. Тимирязева;  
Филатов А.И., к.э.н., профессор, зав. кафедрой экономической кибернетики РГАУ-МСХА им. К.А. Тимирязева;  
Карпузова В.И., к.э.н., доцент, РГАУ-МСХА им. К.А. Тимирязева;  
Скрипченко Э.Н., к.э.н., профессор, РГАУ-МСХА им. К.А. Тимирязева;  
Чернышева К.В., к.э.н., доцент, РГАУ-МСХА им. К.А. Тимирязева.

Материалы публикуются в авторской редакции

Ответственный редактор сборника материалов конференции:  
кандидат технических наук **Н.Б. Паклин**  
e-mail: [edu@basegroup.ru](mailto:edu@basegroup.ru)

## СОДЕРЖАНИЕ

Приветственное слово	5
<b>СЕКЦИЯ «ОПЫТ ПРЕПОДАВАНИЯ ДИСЦИПЛИН С ИСПОЛЬЗОВАНИЕМ АНАЛИТИЧЕСКОЙ ПЛАТФОРМЫ DEDUCTOR»</b>	
<i>Карпузова В.И., Скрипченко Э.Н., Чернышева К.В.</i> Компетентностный подход в обучении магистров по направлению «Менеджмент» профиля «Информационное обеспечение управления АПК» с использованием аналитической платформы Deductor	9
<i>Настащук Н.А.</i> Обучение бакалавров направления подготовки 080100 «Экономика» основам оперативного анализа данных	14
<i>Титов А.П.</i> Проблемы применения аналитических программных комплексов в учебном процессе высшего учебного заведения	19
<i>Красникова С.А.</i> Использование аналитической платформы Deductor на практических занятиях по дисциплине «Методы аналитической обработки данных»	22
<i>Казаков Ю.М., Тищенко А.А., Казаков С.Ю.</i> Опыт использования аналитической платформы Deductor на кафедре «Компьютерные технологии и системы»	26
<i>Мальцева А.В.</i> О перспективах использования современных информационных технологий в практике исследований и обучения в социально-культурной сфере (на примере аналитической платформы Deductor)	30
<i>Шамсутдинова Т.М.</i> Об опыте использования платформы Deductor в учебном процессе Башгосагроуниверситета	31
<b>СЕКЦИЯ «АКТУАЛЬНЫЕ ЗАДАЧИ БИЗНЕС-АНАЛИТИКИ И ИХ РЕШЕНИЕ АЛГОРИТМАМИ DATA MINING»</b>	
<i>Абруков В.С., Троешествова Д.А.</i> Создание баз знаний в задачах прикладной физики	36
<i>Болотова Л.С., Боев Б.В., Иваненко И.Ю.</i> Гибридные системы поддержки принятия решений на основе методов интеллектуального анализа данных	45
<i>Галаев С.А., Потюпкин А.Ю.</i> О некоторых технических приложениях методов Data Mining в области применения и эксплуатации ракетно-космической техники	56
<i>Махныткина О.В.</i> Решение задач оценки компетентности студента с использованием интеллектуального анализа данных	67
<i>Машиинцов Е.А., Найденов А.А.</i> Использование Deductor для оценки продолжительности жизни в отдельном регионе	75

<i>Мороз О.А.</i> Применение методов интеллектуального анализа данных в мониторинге и оценке инновационной активности организаций промышленности	81
<i>Бондарь И.А., Назарова С.А.</i> Информационная модель хранилища данных кафедры вуза	91
<i>Прокопенко Н.Ю., Родионова С.В., Чубченко М.Н.</i> Информационно-аналитическая система мониторинга состояния водных ресурсов Нижегородского региона	100
<i>Рыбанов А.А., Зайчук О.А.</i> Использование АП Deductor для анализа результатов и планирования профориентационной работы вуза	107
<i>Хомич С.В.</i> Некоторые подходы к прогнозированию спроса на продукцию предприятие с использование АП Deductor	109
<i>Татаринов В.В., Скударев М.А.</i> Прогнозирование российской статистики пожаров на 2011 год	114

## ПРИВЕТСТВЕННОЕ СЛОВО

28 июня 2011 года на базе кафедры экономической кибернетики Российского государственного аграрного университета – МСХА им. К.А. Тимирязева (г. Москва) прошла II Межвузовская научно-практическая конференция «Бизнес-аналитика. Использование аналитической платформы Deductor в учебном процессе вуза».

Это ежегодное мероприятие делается как для существующих преподавателей вузов-партнеров BaseGroup Labs, использующих аналитическую платформу *Deductor* в учебном процессе, так и для новых представителей высших учебных заведений, желающих ознакомиться с образовательной инициативой BaseGroup Labs. К конференции проявляют интерес руководители структурных подразделений вузов, заведующие кафедрами, деканы, преподаватели, желающие применять современные информационно-аналитические системы при обучении студентов – будущих прикладных информатиков, специалистов по информационным системам, экономистов, финансистов, математиков. Впервые очный съезд был успешно проведен в июне 2009 года в формате семинара, в 2010 году прошла вторая встреча на базе института информационных и инновационных технологий НОУ «Международная академия бизнеса и управления».

В 2011 году в работе конференции приняли участие 47 участников из 27 вузов России, Украины и Белоруссии, в том числе 18 вузов-партнеров BaseGroup Labs. 28 человек приехали для очного участия в конференции, наибольшее число было из Москвы, но также присутствовали преподаватели из Санкт-Петербурга, Екатеринбурга, Твери и Нижнего Новгорода.



### *Работа конференции*

По тематике присланных докладов и обсуждаемым вопросам можно выделить следующие направления работы конференции.

1. Опыт преподавания дисциплин с использованием аналитической платформы Deductor.

2. Прикладные и исследовательские работы преподавателей и студентов, выполненные на аналитической платформе Deductor.
3. Актуальные задачи бизнес-аналитики и их решение алгоритмами Data Mining.
4. Формирование программ учебных дисциплин, связанных с анализом данных.

Работа конференции началась с приветственного слова организационно-программного комитета: д.э.н., профессор, декан экономического факультета РГАУ-МСХА *Ахметов Р.Г.*, д.э.н., чл.-корр. РАСХН, РГАУ-МСХА им. К.А. Тимирязева *Гатаулин А.М.*, к.э.н., профессор, зав. кафедрой экономической кибернетики РГАУ-МСХА им. К.А. Тимирязева *Филатов А.И.*

Доклад к.э.н., доцента кафедры экономической кибернетики РГАУ-МСХА *Карпузовой Веры Ивановны* был первым. Она рассказала об особенностях преподавания дисциплин, связанных с анализом данных и Data Mining в частности. На кафедре экономической кибернетики применяется идея сквозного обучения студентов: знания подаются последовательно, от простого к сложному. Сначала учащихся знакомят с основами проектирования и работы с OLTP-системами и информационными системами управления предприятием. И только затем студентов знакомят с принципами работы систем поддержки принятия решений, в частности, с аналитической платформой Deductor.

С докладом выступил руководитель образовательных проектов компании BaseGroup Labs, кандидат технических наук *Николай Паклин*. Он рассказал о проблемах учебных программ таких направлений как прикладная информатика и математические методы в экономике. Во многих вузах сильно сокращены часы по предметам, связанным с анализом данных, в связи с переходом на бакалавриат. В докладе также были озвучены мероприятия, прошедшие за последний год, а именно: вышла в свет 2-е издание книги «Бизнес-аналитика», выпущен сборник трудов материалов конференции, проведен конкурс дипломных работ, выбран лучший вуз-партнер 2010 года, выпущено «Руководство по алгоритмам». Выдвинута инициатива проведения олимпиад по бизнес-аналитике.



*Н.Б. Паклин*

Следующий доклад сделал к.т.н., доцент *Андреев В.К.* из Академии Народного Хозяйства и Государствен-

ной Службы (г. Москва). Он также использует в учебном процессе магистрантов аналитическую платформу Deductor, для чего была продемонстрирована программа соответствующей дисциплины. Для успешной сдачи экзамена студентам необходимо в полной мере освоить как теоретические, так и практические материалы, подготовить реферат на одну из тем, в той или иной мере затрагивающей анализ данных, а также выполнить зачетную задачу.

Первая часть конференции завершилась докладом профессора Академии экономической безопасности МВД РФ (г. Москва) *Горбенко А.О.* Аналитическая платформа Deductor в этом вузе используется для обучения студентов-курсантов, будущих профессионалов по борьбе с экономическими преступлениями.

Вторая часть конференции была открыта докладом профессора МГТУ МИРЭА, д.т.н. *Болотовой Л.С.* Она затронула вопрос создания гибридной системы поддержки принятия решения, для которой была использована аналитическая платформа Deductor. В частности, профессор *Болотова Людмила Сергеевна* показала структуру хранилища данных, карты Кохонена с выделенными кластерами городов мира по степени эпидемиологической опасности. Как и в прошлом году, в конце информационного сообщения были озвучены пожелания разработчикам аналитической платформы.

Завершающим докладчиком стал *Белов В.С.*, к.т.н., профессор Института компьютерных технологий Московского государственного университета экономики, статистики и информатики (МЭСИ).

После докладов был проведен круглый стол, на котором преподаватели обменялись мнениями о проблемах преподавания дисциплин по бизнес-аналитике, доступности отечественного программного обеспечения, формой и содержанием проведения олимпиад по бизнес-аналитике, вопросами стыковки бакалаврских и магистерских программ по информационно-аналитическим системам.

Перед закрытием конференции самые активные преподаватели вузов-партнеров уходящего учебного года были награждены грамотами BaseGroup Labs: *Карпузова В.И.*, *Скрипченко Э.Н.*, *Чернышева К.В.*, *Прокопенко Н.Ю.*, *Болотова Л.С.*, *Александрова В.А.*

Компания BaseGroup Labs выражает благодарность руководству экономического факультета Российского государственного аграрного университета – МСХА им. К.А. Тимирязева (г. Москва) и лично доценту, к.э.н. *Карпузовой Вере Ивановне*, профессору *Скрипченко Эльвире Николаевне* за помощь в организации и проведении конференции.

**СЕКЦИЯ**  
**«ОПЫТ ПРЕПОДАВАНИЯ ДИСЦИПЛИН**  
**С ИСПОЛЬЗОВАНИЕМ**  
**АНАЛИТИЧЕСКОЙ ПЛАТФОРМЫ DEDUCTOR»**



# **КОМПЕТЕНТНОСТНЫЙ ПОДХОД В ОБУЧЕНИИ МАГИСТРОВ ПО НАПРАВЛЕНИЮ «МЕНЕДЖМЕНТ» ПРОФИЛЯ «ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ УПРАВЛЕНИЯ АПК» С ИСПОЛЬЗОВАНИЕМ АНАЛИТИЧЕСКОЙ ПЛАТФОРМЫ DEDUCTOR**

*Карпузова В.И., доцент, Скрипченко Э.Н., профессор,  
Чернышева К.В., доцент,  
Российский государственный аграрный университет  
– МСХА имени К.А. Тимирязева, г. Москва*

Компетентностный подход является отличительной чертой Федерального государственного образовательного стандарта высшего профессионального образования (ФГОС ВПО) третьего поколения.

Компетентность – это совокупность компетенций, то есть наличие знаний и опыта, необходимых для эффективной деятельности в определенной предметной области.

Под компетенцией понимается способность применять знания, умения, успешно действовать на основе практического опыта при решении поставленных задач.

Согласно ФГОС ВПО третьего поколения на кафедре экономической кибернетики экономического факультета РГАУ-МСХА имени К.А. Тимирязева открыта магистратура по направлению «Менеджмент» профиля «Информационное обеспечение управления АПК».

Основной образовательной программой подготовки магистров данного профиля предусматривается ориентация выпускников на экспертно-аналитическую, организационно-управленческую, научно-исследовательскую, педагогическую и методическую профессиональную деятельность.

Виды профессиональной деятельности магистра обеспечиваются соответствующими компетенциями, взаимосвязи между которыми приведены в таблице 1.

Компетенции формируются при изучении дисциплин учебного плана подготовки магистра. Учебный план включает общенаучный и профессиональный циклы, которые представлены базовой и вариативной частями.

Дисциплины базовой части носят общепрофессиональный характер. Дисциплины вариативной части определяют профиль подготовки магистра.

Вариативная часть учебного плана подготовки магистров профиля «Информационное обеспечение управления АПК» в университете вклю-

чает следующие дисциплины: «Основы информационного обеспечения управления отраслями АПК»; «Информационные системы в управлении»; «Интеллектуальные системы и технологии»; а также курсы по выбору: «Стандартизация и безопасность информационных систем», «Интернет-технологии в бизнесе», «Базы и хранилища данных информационных систем»; «Прикладной системный анализ», «Математические методы в информационных системах», «Математические методы стратегического планирования».

Таблица 1

Компетенции магистра по направлению «Менеджмент»

Вид профессиональной деятельности	Код компетенции	Наименование компетенции
Организационно-управленческая	ПК-1	Способен управлять организациями, подразделениями, группами (командами) сотрудников, проектами и сетями
	ПК-2	Способен разрабатывать корпоративную стратегию
	ПК-3	Умеет использовать современные методы управления корпоративными финансами для решения стратегических задач
	ПК-4	Способен разрабатывать программы организационного развития и изменений и обеспечивать их реализацию
Аналитическая деятельность	ПК-5	Способен использовать количественные и качественные методы для проведения научных исследований и управления бизнес-процессами
	ПК-6	Владеет методами экономического анализа поведения экономических агентов и рынков в глобальной среде
	ПК-7	Владеет методами стратегического анализа
	ПК-8	Способен готовить аналитические материалы для управления бизнес-процессами и оценки их эффективности
Научно-исследовательская	ПК-9	Способен обобщать и критически оценивать результаты, полученные отечественными и зарубежными исследователями, выявлять и формулировать актуальные научные проблемы
	ПК-10	Способен обосновывать актуальность, теоретическую и практическую значимость избранной темы научного исследования
	ПК-11	Способен проводить самостоятельные исследования в соответствии с разработанной программой
	ПК-12	Способен представлять результаты проведенного исследования в виде научного отчета, статьи или доклада
Педагогическая и методическая деятельность	ПК-13	Способен применять современные методы и методики преподавания управленческих дисциплин
	ПК-14	Способен разрабатывать учебные программы и методическое обеспечение для преподавания управленческих дисциплин

Все дисциплины вариативной части подготовки магистров, как следует из таблицы 2, взаимосвязаны с профессиональными компетенциями.

Для формирования профессиональных компетенций аналитической деятельности в качестве программного обеспечения указанных дисциплин в университете используются системы поддержки принятия управленческих решений, в том числе и Deductor компании BaseGroup Labs.

Таблица 2

Матрица взаимосвязи модульных дисциплин и практик ООП «Информационное обеспечение управления отраслями АПК» с компетенциями магистра по направлению «Менеджмент»

Код компетенции	ПК-1	ПК-4	ПК-5	ПК-6	ПК-7	ПК-8	ПК-9	ПК-11	ПК-12	ПК-13	ПК-14
Основы информационного обеспечения управления АПК		1	1				1		1		
Информационные системы в управлении	1		1		1	1					
Интеллектуальные системы и технологии						1	1		1		
Прикладной системный анализ						1					
Математические методы в информационных системах						1					
Математические методы стратегического планирования		1			1	1					
Стандартизация и безопасность информационных систем											
Интернет-технологии в бизнесе	1			1							
Базы и хранилища данных информационных систем			1					1			

Аналитическая платформа Deductor применяется в университете с 2006 года, а Deductor Studio Professional (профессиональная версия платформы, рабочее место аналитика) – с 2010 года.

Для проведения лабораторных занятий разработаны учебные пособия с реализацией интерактивного способа обучения («сквозная задача») средствами Deductor Studio. Изложенные в пособиях задания обеспечивают выполнение разнообразных функций, включающих сбор данных из различных источников, преобразование и загрузку их в хранилище, хранение информации, получение отчетности, создание произвольных запросов, многомерный анализ и др.

Разбираются разные способы наполнения хранилища данных: путем импорта данных из приложений MS Access и MS Excel, а также загрузки в Deductor Studio Pro условно-постоянной и переменной информации непосредственно из справочников, документов и регистров накопления семантического слоя программного комплекса 1С: Предприятие 8.X.

При использовании Deductor Studio Academic рассматривается способ наполнения хранилища путем импорта данных приложений MS Access и MS Excel, сохраненных в текстовом формате.

Приводится технология работы по визуализации данных с помощью диаграмм, гистограмм, OLAP-кубов, деревьев решений, таблицы сопряженности, самоорганизующейся карты Кохонена и др.

Таким образом, реализуется профессиональная компетенция ПК-5 «Способен использовать количественные и качественные методы для проведения научных исследований и управления бизнес-процессами».

Задания по кластеризации экономических объектов разными методами способствуют формированию компетенции ПК-6 «Владеет методами экономического анализа поведения экономических агентов и рынков в глобальной среде».

Компетенция ПК-7 «Владеет методами стратегического анализа» обеспечивается выполнением заданий по прогнозированию временных рядов, построению деревьев решений и др.

Согласно учебному пособию выполнение заданий оформляется в виде отчетов.

Итак, выполнение всех заданий учебного пособия направлено также на формирование компетенции ПК-8 «Способен готовить аналитические материалы для управления бизнес-процессами и оценки их эффективности».

Следует подчеркнуть, что данные проведенного магистрантами анализа с использованием аналитической платформы Deductor становятся **знаниями** для принятия обоснованного управленческого решения.

В перспективе планируется применение аналитической платформы Deductor в комплексе с системами управления предприятиями и электронного документооборота в учебном процессе и научных исследованиях.

Использование аналитической платформы возможно и при реализации других магистерских программ.

Таким образом, использование в учебном процессе новых информационных и телекоммуникационных технологий, включая аналитическую платформу Deductor, обеспечивает качественное преподавание дисциплин, повышает привлекательность их для

магистрантов, и тем самым реализуется задача подготовки специалистов для современного информационного общества в соответствии с государственным образовательным стандартом третьего поколения.

## **ОБУЧЕНИЕ БАКАЛАВРОВ НАПРАВЛЕНИЯ ПОДГОТОВКИ 080100 «ЭКОНОМИКА» ОСНОВАМ ОПЕРАТИВНОГО АНАЛИЗА ДАННЫХ**

*Настащук Н.А., доцент,*

*Омский государственный университет путей сообщения, г. Омск*

На сегодняшний день применение оперативного анализа данных в экономике и бизнесе обусловлено его эффективностью, быстротой и наглядностью. В этой связи, оперативный анализ данных является неотъемлемой составляющей аналитической деятельности специалиста обозначенной области и должен найти свое место при подготовке будущих экономистов в условиях современного информационного общества.

Рассмотрим один из возможных путей обучения основам оперативного анализа будущих специалистов в области экономики и бизнеса.

Одним из видов профессиональной деятельности бакалавра направления подготовки 080100 «Экономика» является аналитическая деятельность, заключающаяся в осуществлении следующих операций по отношению к обработке и анализу экономической информации:

- поиск информации по полученному заданию, сбор и анализ данных, необходимых для проведения конкретных экономических расчетов;
- обработка массивов экономических данных в соответствии с поставленной задачей, анализ, оценка, интерпретация полученных результатов и обоснование выводов;
- подготовка информационных обзоров, аналитических отчетов.

Кроме того, в федеральном государственном стандарте высшего профессионального образования (ФГОС ВПО) 080100 «Экономика» отмечено [5], что выпускник должен обладать следующими профессиональными компетенциями в области *аналитической деятельности*:

- способностью осуществлять сбор, анализ и обработку данных, необходимых для решения поставленных экономических задач;
- способностью выбрать инструментальные средства для обработки экономических данных в соответствии с поставленной задачей, проанализировать результаты расчетов и обосновать полученные выводы;
- способностью на основе описания экономических процессов и явлений строить стандартные теоретические и эконометрические модели, анализировать и содержательно интерпретировать полученные результаты;
- способностью анализировать и интерпретировать финансовую,

бухгалтерскую и иную информацию, содержащуюся в отчетности предприятий различных форм собственности, организаций, ведомств и использовать полученные сведения для принятия управленческих решений;

- способностью анализировать и интерпретировать данные отечественной и зарубежной статистики о социально-экономических процессах и явлениях, выявлять тенденции изменения социально-экономических показателей;
- способностью, используя отечественные и зарубежные источники информации, собрать необходимые данные проанализировать их и подготовить информационный обзор и/или аналитический отчет;
- способностью использовать для решения аналитических и исследовательских задач современные технические средства и информационные технологии.

На сегодняшний день одним из современных и эффективных информационных средств реализации аналитической деятельности в области экономики и бизнеса является технология OLAP (Online Analytical Processing, оперативный анализ данных), основное назначение которой заключается в проведении оперативной аналитической обработки информации с целью поддержки принятия управленческих решений. Технология OLAP предоставляет конечному бизнес-пользователю средства для формирования и проверки гипотез о свойствах экономических данных или отношениях между ними на основе разнообразных запросов к базе данных [1, 2].

Обучение основам оперативного анализа данных предлагается в рамках дисциплины «Информационные системы и базы данных», которая входит в вариативную часть математического цикла (Б.2) дисциплин основной образовательной программы подготовки бакалавров данного направления. В рамках этого направления подготовки выделяют следующие профили: «Бухгалтерский учёт, анализ и аудит», «Мировая экономика» и «Финансы и кредит». Обучение этому виду анализа средствами OLAP-технологии является весьма актуальным и необходимым для студентов указанных профилей, поскольку большинство современных автоматизированных экономических, в частности бухгалтерских, информационных систем в составе своего функционала имеют возможность построения аналитического отчета с помощью средств OLAP-технологии. Также здесь следует дополнить, что эти системы относят к классу ERP-систем (Enterprise Resource Planning System, система планирования ресурсов предприятия) и получили название Business Intelligence средств [3, 4].

Именно, наличие аналитических компонент в составе функциональных подсистем Business Intelligence средств обеспечивают формирование аналитической отчетности в режиме реального времени, которая в дальнейшем является основой большинства управленческих решений. Приведем некоторые из них: Microsoft Dynamics AX (в прошлом — Ахартa), Microsoft Dynamics NAV (в прошлом — Navision), Парус, БЭСТ и др. Существенным недостатком перечисленных систем является их высокая стоимость и, как следствие этого, невозможность вузов приобрести даже учебную версию подобных программных средств в условиях внедрения концепции перевода учебного процесса вуза на свободное программное обеспечение.

Однако необходимо помнить, что вуз не может обеспечить выпускника знаниями и умениями работать в тех Business Intelligence средств, которые востребованы именно на сегодняшний момент времени, что обусловлено постоянным совершенствованием подобных средств и появлением новых их видов и версий. В этом случае необходимо обеспечить студента такой суммой знаний, умений и навыков, которая позволит ему в дальнейшем, будучи уже на рабочем месте, применять оперативный анализ данных средствами любого средства Business Intelligence.

Одним из вариантов программных средств Business Intelligence является аналитическая платформа Deductor Academic компании BaseGroup Labs – технологическая платформа для создания законченных аналитических решений и предназначена только для образовательных целей. Аналитическая платформа Deductor Academic позволяет осуществить обучение бакалавров направления подготовки 080100 «Экономика» основам оперативного анализа данных, так как реализует суть OLAP-технологии в полном объеме, предоставляя конечному пользователю следующие для этого функциональные возможности:

1. построение хранилища данных;
2. построение гиперкуба<sup>1</sup> (т.е. специально структурированные данные, иначе называемые OLAP-кубами) на основе его структурных элементов – меры и измерения;
3. агрегирование анализируемых данных посредством соответствующих функций;
4. построение единообразной структуры аналитического отчета;
5. единообразная визуализация полученных данных сформированного аналитического отчета.

При этом, хотелось бы отметить следующий методический прием:

---

<sup>1</sup> Гиперкубы предназначены для построения аналитических отчетов, которые могут иметь произвольное сочетание измерения и мер



прежде чем изучать оперативный анализ средствами аналитической платформы Deductor Academic, студентам рекомендуется освоить основы оперативного анализа экономической информации посредством таких информационных технологий обработки числовой информации как «Сводная таблица», «Консолидация» и «Итоги» в рамках общедоступной электронной таблицы MS Excel.

Рассмотрим небольшой практический пример изучения основ оперативного анализа в рамках аналитической платформы Deductor. Для этого предлагается на занятиях использовать базу данных «Борей» (которая входит в состав интегрированного офисного пакета MS Office), содержащую историю продаж за некоторый период и имеющий следующий состав таблиц:

- *Поставщики* – содержит информацию о поставщиках.
- *Клиенты* – содержит информацию о клиентах.
- *Сотрудники* – содержит информацию о сотрудниках фирмы «Борей».
- *Товары* - содержит информацию о товарах.
- *Типы* – содержит информацию о товарной группе товара.
- *Заказано* – содержит информацию о заказе.
- *Заказы* – содержит историю заказов товаров.
- *Доставка* – содержит информацию о доставке.

**Задание:** сформируйте OLAP-куб (т.е. аналитический отчет) согласно варианту задания и ответьте на вопросы, представленные ниже. Вопросы распределены по вариантам в таблице 1.

#### **Список вопросов**

1. Какая товарная группа имеет максимальное количество продаж?
2. Какой товар имеет максимальную сумму продаж?
3. Пять самых продаваемых товаров по выходным (суббота, воскресенье).
4. Семь самых продаваемых товаров.
5. Два самых популярных товаров в каждой товарной группе.
6. Товары, дающие 50% объёма продаж.
7. Товары, дающие 50% прибыли продаж.
8. Постоянные покупатели в каждой товарной группе.

В заключении следует отметить, что обучение бакалавров направления подготовки 080100 «Экономика» основам оперативного анализа данных средствами аналитической платформы Deductor Academic позволяет обеспечить выпускника необходимой суммой знаний, умений и навыков применения OLAP-технологии и в других системах Business Intelligence.

Таблица 1

## Распределение вопросов по вариантам

№ Варианта \ № Вопроса	1	2	3	4	5	6	7	8
1	+	+	+					
2				+	+	+	+	+
3	+			+			+	
4					+	+		+
5		+			+			+
6	+	+	+					
7				+		+	+	
8			+					

**Литература**

1. Андрейчиков А. В. Интеллектуальные информационные системы / А. В. Андрейчиков, О. Н. Андрейчикова. – М.: Финансы и статистика, 2004. – 424 с.
2. Информационные системы и технологии в экономике и управлении: учеб. для студентов вузов / С.-Петерб. гос. ун-т экономики и финансов; под ред. В. В. Трофимова. – М.: Высшее образование, 2009. – 480 с.
3. Насташук Н. А. Некоторые аспекты использования Business Intelligence средств в профессиональной подготовке студентов специальности «Прикладная информатика в экономике» / Н. А. Насташук // Экономика, статистика и информатика. Вестник УМО. – 2008. – № 2. – С. 69-70.
4. Насташук Н. А. Системный подход к использованию Business Intelligence средств в экономических исследованиях / Н. А. Насташук // Информационная среда вуза: материалы XIV Междунар. науч.-техн. конф. – Иваново, 2007. – С. 302-305.
5. Федеральный государственный образовательный стандарт высшего профессионального образования по направлению подготовки 080100 «Экономика» [Электронный ресурс]. Режим доступа: [http://umd.udsu.ru/FGOS\\_VPO/FGOS\\_fail/080100.htm](http://umd.udsu.ru/FGOS_VPO/FGOS_fail/080100.htm).

## **ПРОБЛЕМЫ ПРИМЕНЕНИЯ АНАЛИТИЧЕСКИХ ПРОГРАММНЫХ КОМПЛЕКСОВ В УЧЕБНОМ ПРОЦЕССЕ ВЫСШЕГО УЧЕБНОГО ЗАВЕДЕНИЯ**

*Титов А.П., заведующий кафедрой информационных технологий,  
Московский финансово-юридический университет – МФЮА, г. Москва*

Аналитические программные средства находят все большее применение в организационно-экономической сфере. Опыт практического применения аналитических программных комплексов показывает, что для крупных организаций их применение при решении проблем обеспечения принятия оптимальных решений становится жизненно необходимым. Для предприятий среднего и мелкого бизнеса это зависит от уровня образования их руководителей и менеджеров.

Основные направления в обучении задаются Государственным образовательным стандартом высшего профессионального образования (ГОС ВПО).

В соответствии с ГОС ВПО, при подготовке магистров по направлениям прикладной информатики, информационной бизнес-аналитики, информационного менеджмента должны изучаться информационные технологии анализа бизнес-процессов, интеллектуальные информационные системы и технологии, компьютерные методы статистического анализа и прогнозирования, многокритериальные системы поддержки принятия решений, основы создания и развития информационно-логических, информационно-семантических и информационно-аналитических систем.

Из всех направлений подготовки экономистов, в соответствии с ГОС ВПО, имеется только одна магистерская программа – 521602 «Математические методы анализа экономики», которая предназначена для подготовки высококвалифицированных сотрудников аналитических служб фирм, банков и страховых компаний, экспертных отделов органов государственного управления, научных институтов и высших учебных заведений.

По остальным направлениям подготовки магистров в ГОС ВПО декларируется необходимость изучения аналитики экономических процессов, однако в составе дисциплин нет разделов (дидактических единиц), определяющих необходимость изучения информационных технологий анализа бизнес-процессов с применением соответствующих аналитических программных комплексов.

В Московской финансово-юридической академии аналитические программные комплексы, в том числе и Deductor, применяются для анализа бизнес-процессов при изучении дисциплин «Информационные системы» и «Информационные технологии» (в различных сферах деятельности), «Компьютерное моделирование в экономике», «Интеллектуальные информационные системы», «Бизнес-аналитика», «Автоматизация

экономических расчетов», «Имитационное моделирование процессов», «Макроэкономическая статистика», «Математические методы в экономике», «Математическое моделирование в менеджменте», «Финансовая математика» и другие.

Однако во многих из них изучение методики применения программных комплексов бизнес-аналитики осуществляется в рамках разделов (дидактических единиц), задаваемых ГОС ВПО и поэтому имеет по ряду дисциплин в большей степени ознакомительный характер. Этого явно недостаточно для формирования у студентов навыков в решении задач бизнес-анализа на всех уровнях управления в организационно-экономической сфере.

Отсюда следует, что необходимы новые ГОС ВПО, имеющие отдельные дисциплины или дисциплины с разделами (дидактическими единицами), явно определяющие изучение основ бизнес-анализа построения информационно-логических и математических моделей деятельности предприятий (организаций) и отдельных бизнес-процессов с применением современных аналитических интеллектуальных программных комплексов для студентов всех специальностей.

Анализ ГОС ВПО различных направлений подготовки магистров и бакалавров показывает, что наилучшим образом указанным задачам подготовки магистров, бакалавров и специалистов отвечает дисциплина СД.Ф.03 «Бизнес-аналитика» ГОС ВПО по специальности 080508 – «Информационный менеджмент», квалификация – менеджер, которую целесообразно дополнить некоторыми разделами из дисциплины СД.Ф.02 «Базы данных и базы знаний».

В подобного рода дисциплине четко должны быть определены разделы (дидактические единицы), регламентирующие изучение систем поддержки управленческих решений (DSS/BI) и бизнес-аналитики (BA), их функции (идентификацию, моделирование, прогнозирование, оптимизацию решений, анализ чувствительности и др.), методы и технологии бизнес-аналитики (OLAP -, DM-технологии и др.), системы визуализации данных и решений, генераторы отчетов, методики обнаружения нового знания в хранилищах данных (KDD), модели поиска нового знания, регрессии, прогнозирования временных рядов, кластеризации, ассоциации, последовательности, систем управления моделями (MMS) и знаниями (KMS), аналитических приложений в корпоративных информационных системах, развитие систем бизнес-аналитики, платформы бизнес-интеллекта (Business Intelligence, BI), оценку эффективности систем бизнес-аналитики, управление эффективностью бизнеса (BPM) (компоненты, стандарты), методология проектирования баз данных и знаний, математических моделей, системы и технологии искусственного интеллекта, инструментальные средства и технологии их построения, информационные технологии поддержки принятия решений и управления – в целом, с применением аналитических интеллектуальных программных комплексов, математических моделей и систем искусственного интеллекта.

Включение в состав ГОС ВПО подготовки специалистов, бакалавров и магистров всех сфер деятельности дисциплины, реализующей в комплексе весь спектр информационных технологий поддержки принятия решений, с выделением на её изучение в течение учебного года не менее 148 часов (за счет ликвидации множества дублирующих дисциплин) и современной технической базы, позволит решить проблему обеспечения принятия оптимальных решений в организационно-экономической сфере.

С данной точки зрения аналитические программные комплексы должны совершенствоваться в направлении увеличения количества параметров многомерного анализа и уровня адекватности математического аппарата реальным, многосторонним бизнес-процессам для повышения точности и долгосрочности прогнозирования развития процессов и ситуаций, и их оптимизации в интересах предприятий и организаций.

Аналитические программные комплексы не должны развиваться самостоятельно, а создаваться с расчетом на слияние с другими интеллектуальными средствами, имитационными математическими моделями, многокритериальными системами поддержки принятия решений, системами искусственного интеллекта и др.

Только в этом случае возможно образование интегрированных комплексов поддержки принятия оптимальных решений, причем как в интересах достижения стратегических целей, так и в интересах своевременного, достаточно быстрого и адекватного реагирования на изменения внешних условий в процессе управления деятельностью предприятий и организаций.

# **ИСПОЛЬЗОВАНИЕ АНАЛИТИЧЕСКОЙ ПЛАТФОРМЫ DEDUSTOR НА ПРАКТИЧЕСКИХ ЗАНЯТИЯХ ПО ДИСЦИПЛИНЕ «МЕТОДЫ АНАЛИТИЧЕСКОЙ ОБРАБОТКИ ДАННЫХ»**

*Красникова С.А., старший преподаватель,  
Национальный исследовательский ядерный университет «МИФИ»,  
г. Москва*

В современных условиях процесс формирования компетенций будущих инженеров по специальности «Автоматизированные системы обработки информации и управления» невозможно себе представить без изучения студентами возможностей современных программных продуктов. Наряду с теоретическими знаниями у выпускников должны формироваться навыки использования, как специализированных систем обработки данных, так и универсальных платформ, применимых для анализа данных различного происхождения.

В программу подготовки по специальности «Автоматизированные системы обработки информации и управления» входит изучение дисциплины «Методы аналитической обработки данных». Цель данного курса – дать студентам целостное представление о современных методах, технологиях и инструментальных средствах аналитической обработки данных. В курсе рассматриваются вопросы обработки данных, представленные тремя направлениями: построение запросов и отчетов на детализированных данных, формирование и анализ агрегированных данных, поиск закономерностей в «сырых» данных.

Первый раздел, рассматриваемый в рамках курса, знакомит студентов с основными понятиями изучаемой области знаний, обозначается роль и место методов и технологий аналитической обработки данных в информационном обеспечении различных предметных областей.

Второй раздел посвящен изучению сферы детализированных данных и инструментальных средств создания запросов и генерации отчетов Querying&Reporting. В рамках данного раздела рассматриваются такие вопросы как: поиск информации, реляционные СУБД и язык манипулирования реляционными данными, информационно-поисковые системы, средства доступа к данным и генераторы отчетов, встроенные средства обработки данных генераторов отчетов, словари данных.

В третьем разделе рассматривается сфера агрегированных данных и OLAP-системы. Студенты получают представление о концептуальном многомерном представлении информации, изучают архитектуру OLAP-систем, знакомятся со специальными многомерными СУБД, расширяют свои знания по реляционные технологии и их применимости в многомерном анализе. В результате изучения раздела у студентов формируется комплексный взгляд на собранную в хранилище данных информацию и возможность ее обработки и анализа с использованием инструментальных средств аналитических платформ.

Четвертый раздел ориентирован на изучение сферы поиска закономерностей в данных и модели и методы Data Mining. Изучение данного раздела связано с такими вопросами как: методы интеллектуального анализа данных, поиск функциональных и логических закономерностей в накопленной информации, построение моделей и правил, прогнозирование, стандарты Data Mining, тенденции в области построения систем аналитической обработки данных.

Теория каждого раздела дисциплины подкреплена практическими занятиями в виде лабораторных работ, выполняя которые студенты изучают реализацию определенных методов обработки данных в инструментальных средах. Так, для закрепления теоретического материала третьего и четвертого разделов курса (OLAP и Data Mining) на лабораторных работах применяется отечественная аналитическая платформа Deductor компании BaseGroup Labs.

Выбор данного инструментального средства для поддержки практических занятий курса был обусловлен следующими факторами:

- функциональность платформы Deductor полностью удовлетворяет поставленным задачам курса: на базе данного инструмента могут быть изучены как концепции представления данных в OLAP-системах, так реализация математических методов, способных выявить в исследуемых данных закономерности различных видов (ассоциации, классификации, кластеризации);
- интуитивно понятный интерфейс и русскоязычная справка по системе обеспечивает быстрый старт при использовании платформы Deductor на лабораторных работах;
- наличие готовых тестовых данных и демонстрационных примеров способствует быстрому освоению основных приемов работы в среде Deductor;
- наглядность представления результатов анализа с использованием различных способов визуализации помогает студентам интерпретировать полученные закономерности даже для малоизвестных предметных областей;
- наличием бесплатной образовательной версии системы для использования в учебных целях – Deductor Academic.

Все вышеперечисленные факторы делают инструмент весьма привлекательным для его применения в учебном процессе.

Лабораторный практикум включает несколько блоков лабораторных работ, каждый из которых связан с изучением определенного аспекта анализа данных:

1. Исследование агрегированных данных;
2. Ассоциативные правила;
3. Классификация данных;
4. Кластеризация данных.

В свою очередь каждый блок включает изучение определенных

математических методов и их реализацию в аналитической платформе Deductor. В первом блоке, «Исследование агрегированных данных», изучаются базовые приемы работы с инструментом «Импорт данных», а также универсальные методы обработки: очистка и трансформация данных, способы визуализации данных.

Практическое задание каждой лабораторной работы состоит из двух разделов. Первый раздел предполагает выполнение студентом пошагового сценария работы с инструментом, а второй раздел содержит индивидуальное творческое задание.

Сценарий первого блока практических заданий при выполнении лабораторной работы блока «Исследование агрегированных данных» выглядит следующим образом.

Студентам предлагается подготовить данные для анализа и исследовать их:

- для данных из файла/файлов (например, Credit.txt) настроить 5 кубов для OLAP-анализа данных в соответствии с представленными ниже описаниями;
- для визуализации результатов анализа использовать табличное (куб) и графическое представления (кросс-диаграмма).

Приведем описание нескольких отчетов.

1. Отчет для анализа зависимости между суммой кредита, целью кредита, полом клиента и его уровнем образования. Измерения: цель кредита, пол клиента, уровень образования. Факт: сумма кредита.

2. Отчет, отображающий зависимость между размером кредита (преобразовать непрерывные данные в дискретные), гражданским состоянием, наличием квартиры, загородного дома, земельного участка и решением о выдаче кредита. Измерения: размер кредита, гражданское состояние, наличие квартиры, загородного дома, земельного участка. Факт: решение о выдаче кредита.

При формировании данных отчетов студент осваивает сценарии работы в среде Deductor и его основные возможности.

Далее студент получает индивидуальное творческое задание, выполняя которое он закрепляет полученные знания и практические навыки. Студенту предлагается для данных из файла (файл и его описание выдается преподавателем) описать и настроить 5 кубов для OLAP-анализа. Описание должно включать: название куба; краткое описание назначения (что с его помощью будет проанализировано); перечень измерений; перечень мер (фактов); для визуализации результатов анализа используется табличное (куб) и графическое представления (кросс-диаграмма).

На защите лабораторной работы студент должен ответить на теоретические вопросы, а также проинтерпретировать полученные результаты, как по общей части работы, так и по индивидуальному заданию.

Таким образом, при выполнении лабораторного практикума на базе платформы Deductor студенты закрепляют теоретические знания по



методам аналитической обработки данных и приобретают навыки применения современного программного продукта, поддерживающего данные методы.

## **ОПЫТ ИСПОЛЬЗОВАНИЯ АНАЛИТИЧЕСКОЙ ПЛАТФОРМЫ DEDUCTOR НА КАФЕДРЕ «КОМПЬЮТЕРНЫЕ ТЕХНОЛОГИИ И СИСТЕМЫ»**

*Казаков Ю.М., доцент, Тищенко А.А., доцент,  
Казаков С. Ю., аспирант,*

*Брянский государственный технический университет, г. Брянск*

В настоящее время информатизация всех отраслей народного хозяйства и развитие информационных систем является одним из прогрессивных и перспективных направлений в науке и технике. В связи с этим непрерывно растет потребность в квалифицированных специалистах, способных успешно работать в области науки и техники, включающей в себя совокупность средств, способов и методов человеческой деятельности, направленных на создание и применение систем сбора, передачи, обработки, хранения и накопления информации.

Кафедра «Компьютерные технологии и системы» основана в Брянском государственном техническом университете в 2003 году и ведет подготовку по 4 направлениям: «Информатика и вычислительная техника», профиль «Системы автоматизированного проектирования»; «Информационные системы и технологии»; «Информационная безопасность автоматизированных систем»; «Информационная безопасность», профиль «Организация и технология защиты информации». Также имеется и магистерская подготовка по данным направлениям.

При подготовке специалистов большое место занимает получение ими знаний по сбору и анализу различных данных, использованию математических методов и подходов к принятию решений в сфере своей профессиональной деятельности. Процессы информатизации, затронувшие различные отрасли деятельности, рост объема информации привели к тому, что для принятия адекватного, обоснованного анализа специалистам во всех областях, а особенно в области разработки, внедрения и защиты информационных технологий, требуется сбор и обработка большого количества информации.

Для решения данных задач достаточно эффективно применяются информационно-аналитические системы, которые позволяют хранить, обрабатывать и анализировать большие объемы данных, поддерживающие оба направления анализа данных: оперативный и интеллектуальный анализы данных.

Оперативный (OLAP) анализ позволяет извлекать из всего объема данных информацию необходимую конкретному лицу, по конкретному вопросу, за короткий промежуток работы информационно-аналитической системы.

Интеллектуальный анализ данных (Data Mining) использует современный математический аппарат, например нейронные сети, генетические алгоритмы, нечеткая логика, и позволяет выявлять в наборе данных

неочевидные закономерности.

В настоящее время большую актуальность при подготовке специалистов, особенно в области информатизации, приобретает, с одной стороны, теоретическое и практическое овладение современным математическим аппаратом, как оперативного, так и интеллектуального анализа данных, а с другой – умение работать с информационно-аналитическими системами или платформами, позволяющих автоматизировать данные анализы.

Для решения таких задач на кафедре «Компьютерные технологии и системы» используют следующие подходы к обучению.

1. Включение в учебный план подготовки студентов полноценных курсов лекций по сбору и анализу данных, а именно: *Модели и методы интеллектуального анализа данных; Интеллектуальные информационные систем;*, *Нейросетевое информационное моделирование; Модели и методы анализа проектных решений; Методы принятия решений.*
2. Внедрение в учебные дисциплины лабораторных и практических занятий, в ходе которых на основе сбора и анализа экспериментальных данных решаются конкретные задачи в определенной предметной области: *Управление данными; Промышленная логистика; Организация и управление службой защиты информации; Средства и системы технического обеспечения обработки, хранения и передачи информации; Банки и базы данных, Компьютерные технологии в инженерном деле; Управление в социальных и экономических системах*
3. Знакомство с имеющимися на российском рынке информационно-аналитическими системами.
4. Проведение анализа данных с использованием аналитической платформы Deductor.
5. Самостоятельное использование системы Deductor при решении практических задач, встающих в ходе выполнения курсовых и дипломных работ. Например: *Курсовая работа по теме «Создание искусственной нейронной сети (самоорганизующаяся карта Кохонена) для прогнозирования котировок акций Сбербанка и ВТБ»* (выполнена магистрантом Ю.Ю. Малаховой).
6. Научно-исследовательская работа со студентами.

Для успешного обучения важно сочетать теоретические знания с практической реализацией знаний при решении конкретных задач. Анализ данных сегодня невозможно проводить без специализированных пакетов программ, реализующих хранение данных, их структуризацию, визуализацию, математический аппарат для выявления закономерностей, программного инструмента, реализующего соответствующий математический метод. На российском рынке имеется большое количество аналитических систем, которые в той или иной степени позволяют анализировать данные с использованием как традиционных статистических мето-

дов, так и современных математических инструментов интеллектуального анализа данных. Но как показал обзор имеющихся систем, часть из них дорогие или тяжелые для восприятия студентами, часть не имеют комплексного подхода к анализу данных, и использует только какой-то определенный математический аппарат, например, пакет Statistica использует только традиционный статистический анализ.

В итоге в качестве универсальной среды для анализа данных была выбрана отечественная разработка – Deductor. Deductor является аналитической платформой, т.е. основой для создания законченных прикладных решений. Реализованные в Deductor технологии позволяют на базе единой архитектуры пройти все этапы построения аналитической системы: от создания хранилища данных до автоматического подбора моделей и визуализации полученных результатов.

Deductor предоставляет аналитикам инструментальные средства, необходимые для решения самых разнообразных аналитических задач: корпоративная отчетность, прогнозирование, сегментация, поиск закономерностей – эти и другие задачи, где применяются такие методики анализа, как OLAP и Data Mining. Deductor является идеальной платформой для создания систем поддержки принятий решений.

Реализованная в Deductor архитектура позволяет добиться максимальной гибкости при создании законченного решения. Благодаря данной архитектуре можно собрать в одном аналитическом приложении все необходимые инструменты анализа и реализовать автоматическое выполнение подготовленного сценария.

Технологическая платформа включает средства, позволяющие максимально сократить сроки разработки, быстро создавать и выводить на рынок новые прикладные решения и быстро адаптировать их в соответствии с изменяющимися требованиями учебного процесса.

Для создания законченного решения достаточно получить данные, определить сценарий обработки и задать место для экспорта полученных результатов. Наличие мощного набора механизмов обработки и визуализации позволяет двигаться по шагам, от наиболее простых способов анализа к более мощным.

Deductor объединяет все необходимые для анализа инструменты, представляет пользователям огромные возможности:

- мощная аналитическая платформа;
- современные самообучающиеся механизмы анализа;
- единое хранилище данных;
- единый пользовательский интерфейс для любых механизмов анализа;
- пакетное выполнение сценариев обработки;
- удаленная аналитическая обработка;
- отделение работы аналитика от конечного пользователя.

Кроме рассмотренных выше возможностей плюсом аналитической платформы Deductor является наличие различных методических мате-

риалов, объясняющих как теоретические понятия и подходы к анализу данных, так и механизмы построения практических решений в аналитической платформе, а также бесплатно распространяемая учебная версия и возможность обучения преподавательского состава. Это позволяет студентам освоить в должной мере данный программный продукт.

Практика преподавания теоретических основ анализа данных и решения конкретных задач с помощью аналитической системы Deductor на кафедре «Компьютерные технологии и системы» выявило ряд проблем, которые следует учитывать при подготовке будущих специалистов по выпускаемым кафедрой направлениям.

Во-первых, следует уделять особое внимание базовым дисциплинам, таким как общий курс математики, теория вероятностей и математической статистики, так как они являются фундаментом для изучения интеллектуального анализа данных. Пробелы в данных курсах могут стать отталкивающим фактором при начале обучения и практическом применении аналитической платформы Deductor.

Во-вторых, аналитическая платформа не решает задачи аналитика, а является всего лишь мощным программным инструментом, помогающим решать практические задачи, поэтому для правильной постановки задачи и интерпретации необходимы математические знания и навыки компьютерной реализации различных результатов анализа данных, и возможности по использованию того или иного метода при решении конкретной задачи.

В заключении подчеркнем, что преподавателям необходимо мотивировать студентов к изучению современных методов и инструментов анализа данных в качестве конкурентного преимущества при поиске работы по специальности в будущем, несмотря на некоторую сложность теоретических и прикладных аспектов данного направления. А для более детальной проработки знаний и практического использования методов и программной реализации следует активнее внедрять в курсовое и дипломное проектирование решение конкретных практических задач с использованием аналитической платформы Deductor.

Несмотря на данные проблемы, опыт преподавания показывает, что студенты, мотивированные на получения современных знаний в области информационных технологий и инструментов проведения перспективного интеллектуального анализа данных готовы обучаться, использовать все возможности платформы Deductor и творчески подходить к решению различных задач.

# **О ПЕРСПЕКТИВАХ ИСПОЛЬЗОВАНИЯ СОВРЕМЕННЫХ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ В ПРАКТИКЕ ИССЛЕДОВАНИЙ И ОБУЧЕНИЯ В СОЦИАЛЬНО-КУЛЬТУРНОЙ СФЕРЕ (НА ПРИМЕРЕ АНАЛИТИЧЕСКОЙ ПЛАТФОРМЫ DEDUCTOR)**

*Мальцева А.В., доцент,  
Алтайский государственный университет, г. Барнаул*

Современный этап развития научного знания характеризуется активным внедрением информационных технологий. Для естественных и технических наук очевидно даже деление этих технологий на традиционные и современные, а в некоторых случаях на новейшие. Науки гуманитарные таким богатством подходов в этом смысле не отличаются, но все же неуклонно двигаются в этом направлении. Актуальным представляется рассмотрение примеров использования некоторых современных информационных технологий для решения прикладных задач социально-культурной сферы.

Диапазон определений понятий «технология» и «информационная технология» достаточно широк, остановимся на тех из них, которые используются в литературе по социологическим и политическим исследованиям. Например, «технология можно определить как объединенную в общую систему и обеспечивающую большую эффективность той или иной деятельности совокупность разнородных компонентов: знаний, методов, операций и правил, а также энергетических сырьевых, технических, кадровых и прочих ресурсов» [1, с. 14] или: «Информационная технология – это совокупность знаний о способах автоматизированной переработки информации с использованием ЭВМ для автоматизации... деятельности» [2, с. 11]. По функциональным признакам, для обеспечения исследовательской деятельности в социально-культурной сфере выделяют информационно-аналитические и информационно-прогнозные технологии [3, с. 39]. Остановимся подробнее на первой разновидности информационных технологий применительно к конкретному примеру программного продукта. Очевидно, что первоочередной целью являются сбор, обработка и анализ информации об исследуемом объекте, с возможностью принятия взвешенного решения.

Эволюция технологий в этой области давно шагнула за пределы привычных пакетов статистической обработки данных, таких как SPSS, AcaStat, RealStat, Да-система, Статистика и др. Конец XX в. ознаменовался стремительным ростом объемов информации, накопленной на различных носителях, соответственно возросли потребности научного и бизнес-сообществ по использованию различных комбинаций таких данных в сочетании в том числе и с имеющимися архивными электронными базами данных. Объемы роста цифровой информации, прогнозируемые

Агентством IDC к 2010 г., составят тысячи экзабайт (1 экзабайт = 260 байт, или миллиард гигабайт) [4]. Появились аналитические платформы, позволяющие создавать комплексные прикладные решения (еще их называют законченными прикладными решениями) в области анализа данных. Их преимуществом является возможность использования наборов интегрированных между собой инструментов бизнес-анализа [5]. В аналитических платформах успешно реализуются новые парадигмы анализа больших объемов данных: Data Mining (интеллектуальный анализ данных), Data Warehousing (хранилища данных), Knowledge Discovery in Databases (извлечение знаний из баз данных) [6]. Все перечисленные способы анализа объединяются общим принципом обнаружения ранее неизвестных, нетривиальных, практически полезных и доступных интерпретаций знаний.

В настоящее время развитие этой области информационных технологий отличается высокой конкурентной средой, поскольку разработка ведется несколькими лабораториями, представляющими на рынок собственные продукты. К наиболее известным производителям аналитических платформ относятся BaseGroup Labs и их продукт Deductor, существующий в линейке лицензий от Academic до Enterprise и Professional; Business Objects с одноименным продуктом; IBM с Cognos; SAS с SAS Enterprise Miner и Forecast Server; Oracle с Oracle BI.

Применение аналитических платформ в практике анализа данных в социально-культурной сфере перспективно. При подготовке специалистов данного профиля на факультете социологии АлтГУ кафедрой математических методов в социальных науках используется продукт Deductor Academic компании BaseGroup Labs.

### **Литература**

1. Яковлев, И.Г. Информационно-аналитические технологии и политическое консультирование/ И.Г. Яковлев// Полис. – 1998. – №2 – 3.
2. Уткин, В.Б. Информационные системы и технологии в экономике/ В.Б. Уткин, К.В. Балдин. - М., 2005.
3. Боришполец, К.П. Методы политических исследований: Учебное пособие/К.П. Боришполец. – М., 2005.
4. Левкович-Маслюк, Л. Великие раскопки и великие вызовы/ Л. Левкович-Маслюк // Компьютера. [Электронный ресурс]. – Режим доступа: <http://www.computerra.ru/features/311656/>
5. Резниченко А. Аналитические платформы в «магическом квадрате» Gartner. [Электронный ресурс]. – Режим доступа: [www.docflow.ru/analytic](http://www.docflow.ru/analytic)
6. [www.basegroup.ru](http://www.basegroup.ru)

## ОБ ОПЫТЕ ИСПОЛЬЗОВАНИЯ ПЛАТФОРМЫ DEDUCTOR В УЧЕБНОМ ПРОЦЕССЕ БАШГОСАГРОУНИВЕРСИТЕТА

*Шамсутдинова Т.М., доцент,  
Башкирский государственный аграрный университет, г. Уфа*

Кафедрой информатики и информационных технологий Башкирского государственного аграрного университета проводится ряд занятий, связанных с обучением студентов навыкам интеллектуального анализа данных. В частности, такие занятия предусмотрены в курсе «Интеллектуальные информационные системы» на специальности 080801 *Прикладная информатика (в экономике)* и для направления подготовки бакалавров 230700 *Прикладная информатика*.

При этом для развития у студентов умения и навыков извлечения знаний кафедрой предусмотрены следующие лабораторные работы с использованием Deductor Studio Academic.

1. «Знакомство с системой Deductor Studio. Построение деревьев решений».
2. «Анализ данных с использованием нейронной сети».
3. «Кластеризация данных с использованием самоорганизующихся карт Кохонена».
4. «Использование системы Deductor Studio для комплексного анализа данных».

В лабораторной работе «*Знакомство с системой Deductor Studio. Построение деревьев решений*» рассматривается общее понятие интеллектуальной «добычи» данных Data Mining, описываются возможности платформы Deductor по интеллектуальному анализу данных, изучаются общие принципы создания и применения деревьев решений.

В ходе выполнения работы студентам предлагается ознакомиться с демопримерами интеллектуального анализа данных в системе Deductor Studio, проанализировать пример построения дерева решений в данной системе, рассмотреть, как меняется число правил в зависимости от разного процента отсечения ветвей дерева.

Далее студентам предлагается выполнить задания для самостоятельной работы, при этом провести анализ полученных данных, выявить наиболее значимые атрибуты для принятия решения.

В работе «*Анализ данных с использованием нейронной сети*» даются понятия искусственного нейрона и нейронных сетей, приводятся принципы построения нейросетей средствами системы Deductor Studio. Далее рассматривается пример использования нейронных сетей при решении задачи оптимизации. В ходе реализации данного примера студенты должны подготовить файл с исходными данными и импортировать его в систему. Далее с использованием мастера обработки им нужно создать пользовательскую модель для расчета общей прибыли торговой ор-



ганизации. Для нахождения оптимальной стратегии продаж, обеспечивающей наибольшую прибыль, необходимо использовать узел «Нейросеть». На основе построенной нейросетевой модели студентам предлагается провести интерпретацию результатов, позволяющую сделать заключение об оптимальной стратегии продаж для торговой организации.

Работа «*Кластеризация данных с использованием самоорганизующихся карт Кохонена*» включает рассмотрение таких вопросов, как понятие кластеризации данных, принципы построения самоорганизующихся карт Кохонена в системе Deductor. При этом дается пример применения кластерного анализа на основе самоорганизующихся карт Кохонена и предлагаются задания для самостоятельного выполнения. При этом необходимо провести анализ полученных данных, выявить взаимосвязи между наиболее значимыми атрибутами для принятия решения.

В лабораторной работе «*Использование системы Deductor Studio для комплексного анализа данных*» рассматривается пример анализа экономической информации. Комплексный анализ данных при этом включает в себя кластерный анализ (с использованием самоорганизующихся карт Кохонена), построение нейронной сети (с использованием узла «Нейросеть») и построение дерева решений (при помощи узла «Дерево решений»).

Студентам предлагается реализовать следующие этапы анализа данных:

- на основании открытой статистической отчетности сформировать набор данных, содержащий ряд основных социально-экономических показателей для заданных экономических объектов;
- с использованием инструмента анализа «Карта Кохонена» системы Deductor Studio Academic провести кластерный анализ данных, в результате которого выявить группы объектов с различным уровнем социально-экономических показателей;
- с использованием инструмента анализа «Нейросеть» построить нейронную сеть, позволяющую проанализировать влияние отдельных факторов на отнесение объекта к тому или иному кластеру. При обучении нейронной сети необходимо использовать результаты кластерного анализа данных в виде самоорганизующихся карт Кохонена;
- построить дерево решений, показывающее условия принадлежности объекта к определенной группе. В результате анализа данных выявить наиболее значимые факторы для принятия решения об отнесении объекта к определенному кластеру.

Для закрепления знаний по работе с системой Deductor Studio студентам предлагаются следующие вопросы для самоконтроля знаний:

- Что понимают под термином Data mining?
- Приведите примеры технологий, относящихся к методам Data

mining.

- Какие средства интеллектуального анализа экономической информации реализованы в системе Deductor Studio?
- Охарактеризуйте понятие деревьев решений.
- Что представляет собой правило дерева решений?
- Из каких этапов состоит процесс построения дерева решений в Deductor Studio?
- Охарактеризуйте понятие искусственного нейрона.
- Что понимают под обучением нейронной сети?
- Приведите примеры задач, решаемых нейросетевыми методами.
- Как вызвать инструмент анализа «Нейросеть» Мастера обработки системы Deductor?
- Из каких шагов состоит процесс построения и обучения нейросети в системе Deductor?
- Перечислите основные способы отображения нейронной сети в системе Deductor.
- Что понимают под кластеризацией данных?
- Приведите примеры задач, решаемых с использованием кластерного анализа.
- Что представляет собой алгоритм функционирования самоорганизующихся карт Кохонена?
- В чем заключается процесс обучения самоорганизующейся карты?
- Как вызвать инструмента анализа «Карта Кохонена» в системе Deductor Studio?
- Перечислите основные этапы построения самоорганизующейся карты в системе Deductor Studio.

В курсе «Интеллектуальные информационные системы» студенты также выполняют курсовую работу, связанную с проектированием базы знаний экспертной системы. Для извлечения знаний из данных им предлагается использовать методы интеллектуального анализа данных, в том числе и с применением Deductor Studio.

Таким образом, можно заключить, что платформа Deductor предоставляет широкие возможности по обучению студентов навыкам интеллектуального анализа данных. Данная платформа может быть использована и в ходе аудиторной работы, и при организации самостоятельной работы студентов.

**СЕКЦИЯ**  
**«АКТУАЛЬНЫЕ ЗАДАЧИ**  
**БИЗНЕС-АНАЛИТИКИ И ИХ РЕШЕНИЕ**  
**АЛГОРИТМАМИ DATA MINING»**

## **СОЗДАНИЕ БАЗ ЗНАНИЙ В ЗАДАЧАХ ПРИКЛАДНОЙ ФИЗИКИ**

*Абруков В.С., профессор, Троешестова Д.А., доцент,  
Чувашский государственный университет имени И.Н.Ульянова,  
г. Чебоксары*

**Постановка задачи.** Во многих научных и научно-производственных организациях России, а также за рубежом, долгое время занимавшихся исследованием различных физических процессов, накоплено много экспериментальных данных. Перед исследователями возникает вопрос: можно ли их обобщить и на этой основе представить в виде многофакторных количественных моделей, позволяющих обобщать ранее полученные экспериментальные результаты, предсказывать на этой основе закономерности явлений ранее не исследованных систем или для ранее не исследованных условий? Еще более важным вопросом является следующий: какие системы необходимо использовать для реализации требуемых характеристик физического явления в заданных технических условиях? Аналогичные вопросы стоят в области нанотехнологий.

Таким образом, мы поставили задачу, заключающуюся в сборе и обработке экспериментальных данных, представлении их в виде многофакторных вычислительных моделей, позволяющих обобщать ранее полученные экспериментальные данные и предсказывать на этой основе закономерности исследуемых физических явлений ранее не исследованных систем или для ранее не исследованных условий. Объектами исследования были: горение конденсированных систем (КС), вибрационное горение (ВГ) газов, а также новый продукт нанотехнологий – линейно-цепочечный углерод с внедренными в него различными химическими элементами (ЛЦУ).

В качестве методов решения данных задач были выбраны средства Data Mining, в частности искусственные нейронные сети.

Использовались как собственные экспериментальные данные, так и экспериментальные результаты, опубликованные в России.

**Методы исследования.** Data Mining – комплекс современных средств предобработки данных, их анализа и моделирования [1-4]. Data Mining включает в себя такие средства, как: методы очистки данных и подготовки их к анализу, факторный и корреляционный анализ, деревья решений, искусственные нейронные сети (ИНС) [2, 3, 4], самоорганизующиеся карты Кохонена [4]. С точки зрения создания многофакторных вычислительных моделей, ИНС, которые могут рассматриваться как

универсальные аппроксиматоры многомерных функций, играют главную роль. Остальные средства служили нам для предобработки данных и для их предварительного анализа.

Основы применения средств Data Mining приведены в [1-4]. Ряд примеров использования Data Mining в научных и прикладных исследованиях приведены в работах авторов [5-8]. В работе [5] – для построения вычислительных моделей при решении прямых и обратных задач оптики на основе неполных данных об изображении, в частности на основе измерения оптического сигнала всего в одной точке плоскости изображения («одноточечное измерение»); в [6] – для предсказания формы волны на свободной поверхности жидкости (задача цунами); в [7] – для создания моделей социальных явлений; в [8] – для создания модели перехода от медленного горения к детонации при различных условиях эксперимента.

Методики применения средств Data Mining, разработанные авторами, приведены на Интернет-сайте электронной библиотеки Чувашского государственного университета [9] и на Интернет-сайте [10].

**Безгазовые конденсированные системы (самораспространяющийся высокотемпературный синтез): модели «Температурные профили волны горения».** Задача непосредственного определения температурного профиля в волне горения безгазовых КС относится к трудным экспериментальным задачам. В первую очередь это связано с трудностями применения термопар (наиболее распространенный метод определения температуры) при очень высоких температурах и градиентах температуры, реализующихся в волне горения. Эти измерения требуют очень тщательной организации эксперимента, очень трудоемки и дороги.

Полученные нами ИНС-модели позволяют обобщать ранее полученные многомерные данные и на этой основе интерполировать и экстраполировать их, то есть получать интересующие исследователя многомерные данные без проведения экспериментов.

Для построения моделей нами была использована аналитическая платформа Deductor [4].

Схема построения моделей была следующая.

Сначала были оцифрованы температурные профили в волне горения безгазовых КС-смесей порошков Ti, В, Fe и ферроборных сплавов Fe-nВ с титаном, отличающиеся содержанием железа в КС и начальной температурой образца. Затем была выбрана соответствующая собранной базе данных архитектура ИНС и проведено ее обучение (рис. 1). Обучение заключалось в том, что различные наборы значений расстояния от поверхности горения, для различных типов системы, содержания железа в них и начальной температуры подавались на входной слой ИНС, а соответст-

вующие значения температуры устанавливались в выходном слое ИНС и с помощью известного метода обучения ИНС – метода «обратного распространения ошибки» [4] – создавалась вычислительная ИНС модель горения.

Эта модель представляет собой модель типа «чёрного ящика». Полученный «чёрный ящик» может использоваться для определения (прогнозирования) температурного профиля в волне горения безгазовой КС примерно такого же типа, для которого была собрана база данных следующим образом. Тип системы, содержание железа в ней и начальная температура устанавливаются во входном слое ИНС, там же указывается произвольное (или интересующее исследователя) значение расстояния от поверхности горения. После этого «чёрный ящик» мгновенно вычисляет соответствующие значение температуры, а также профиль температуры в целом (автоматически перебирая равномерно распределенные значения расстояния от поверхности горения в пределах значений имеющихся в обучающей базе данных).

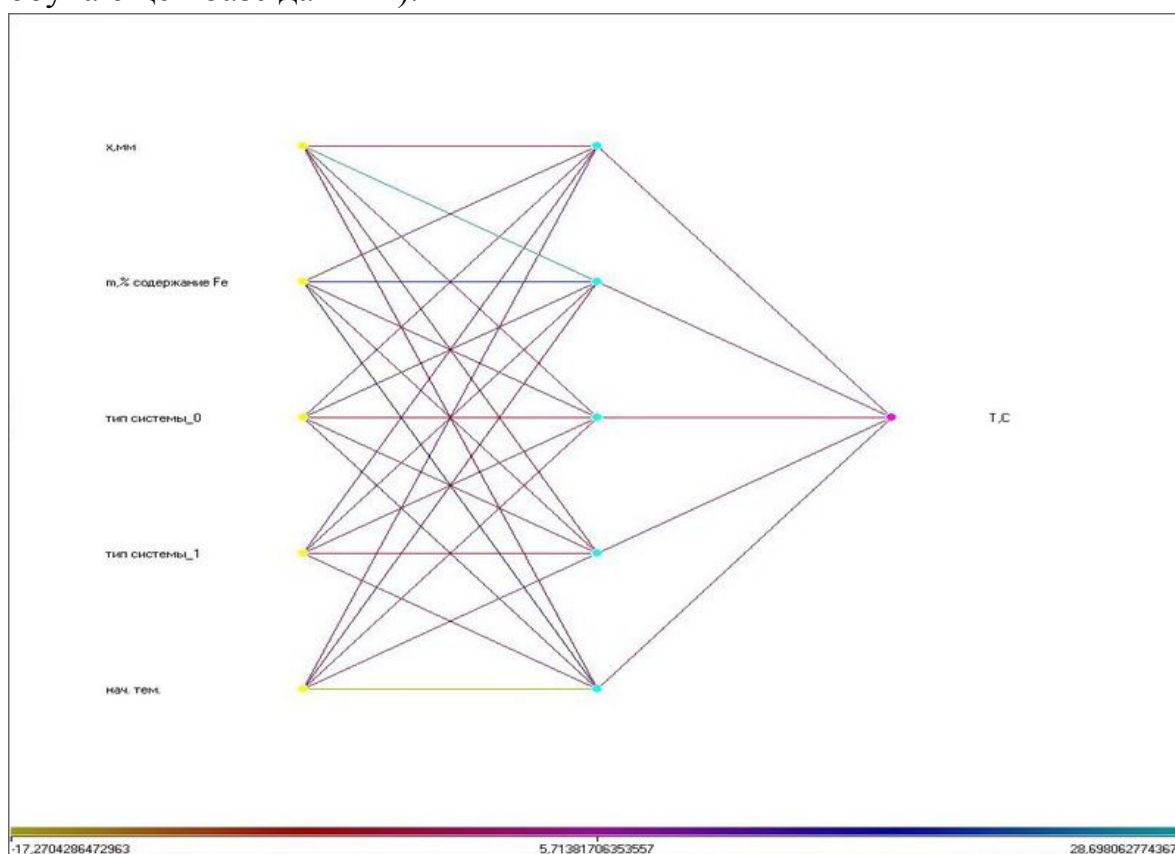


Рис. 1. Структура ИНС, использованная при получении ИНС-модели «Температурные профили волны горения» для безгазовых КС – смесей порошков Ti, В, Fe и ферроборных сплавы Fe-nВ с титаном

Ниже представлены результаты и их обсуждение.

В обучающей базе данных было 217 строк – наборов вышеуказанных величин. База данных была проверена с помощью узла-обработчика

Deductor в группе очистки данных «Выявление дубликатов и противоречий».

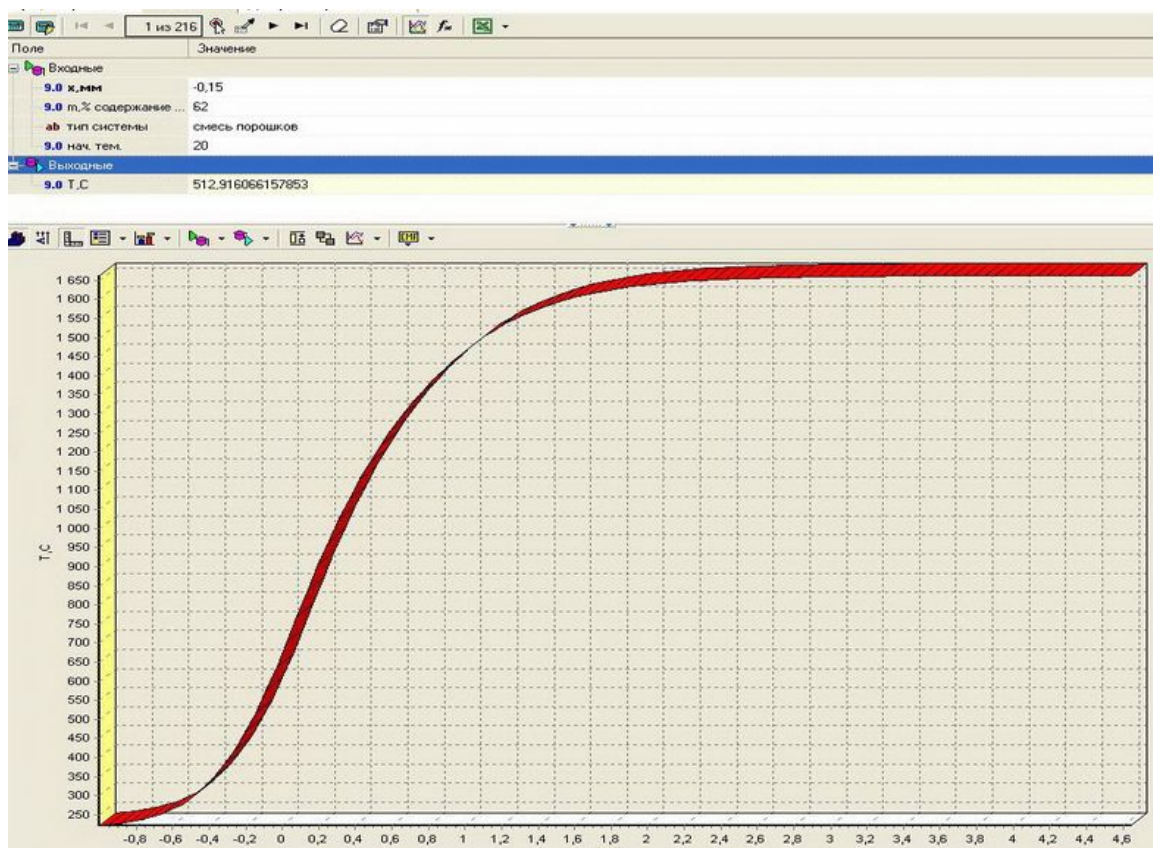


Рис. 2. Профиль температуры в волне горения безгазовой КС – смеси порошков Ti, V, Fe при содержании железа – 62%, начальной температуре 20 °С. Эти значения, расстояние расстояния от поверхности горения x и соответствующее значение температуры (Т) отображены в верхней части рисунка – над графиком

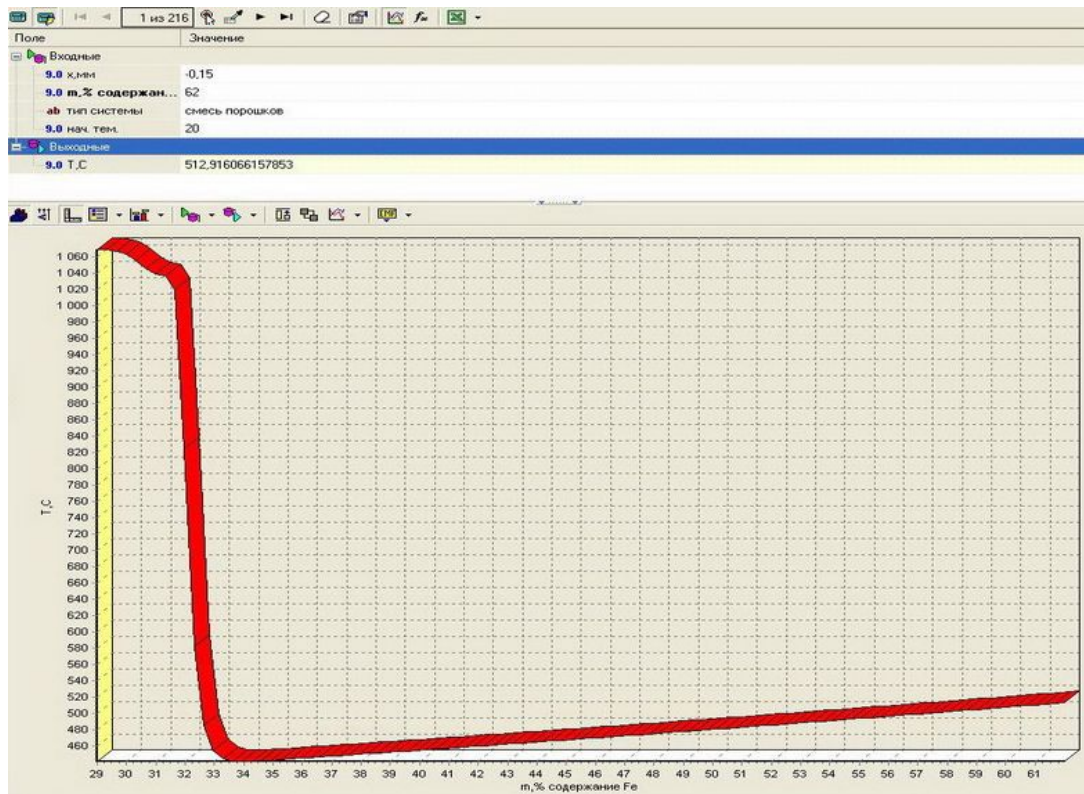


Рис. 3. Зависимость температуры в волне горения (на расстоянии от поверхности горения  $x = -0,15$  мм) от содержания железа (для безгазовой КС – смеси порошков Ti, В, Fe при начальной температуре образца 20 °C)

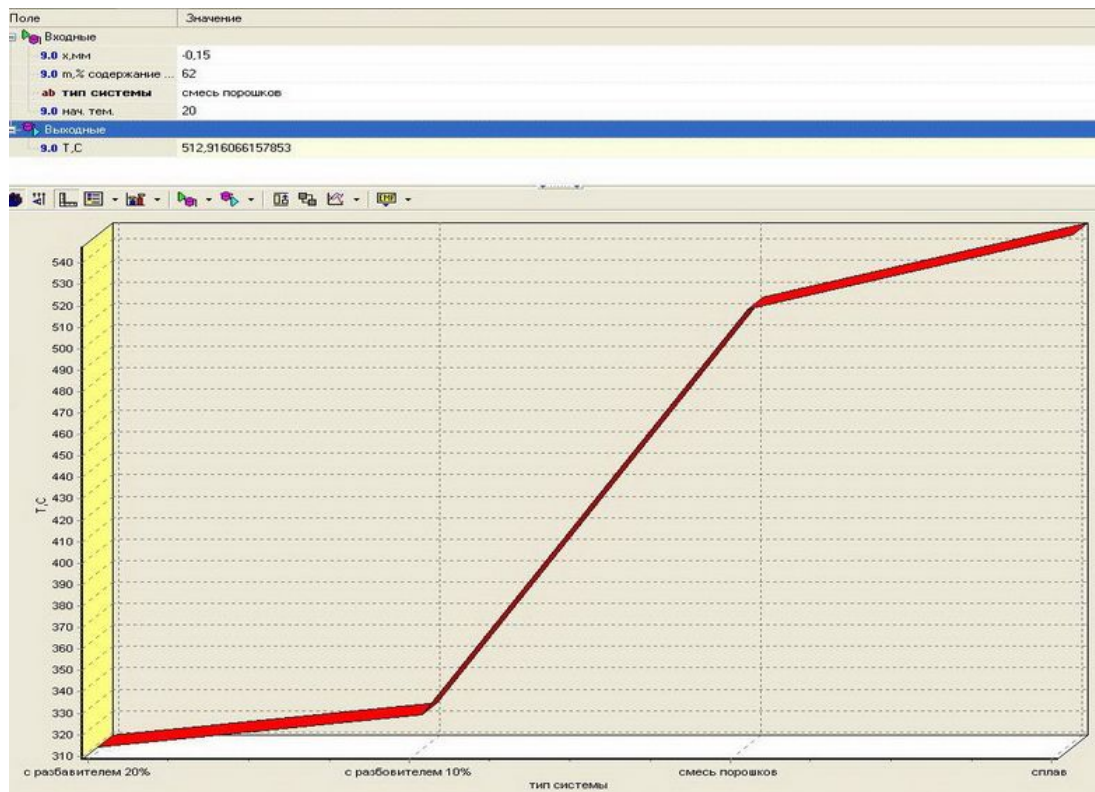


Рис. 4. Зависимость температуры в волне горения (на расстоянии от поверхности горения  $x = -0,15$  мм) от типа системы (для безгазовой КС) при начальной температуре образца 20 °C)



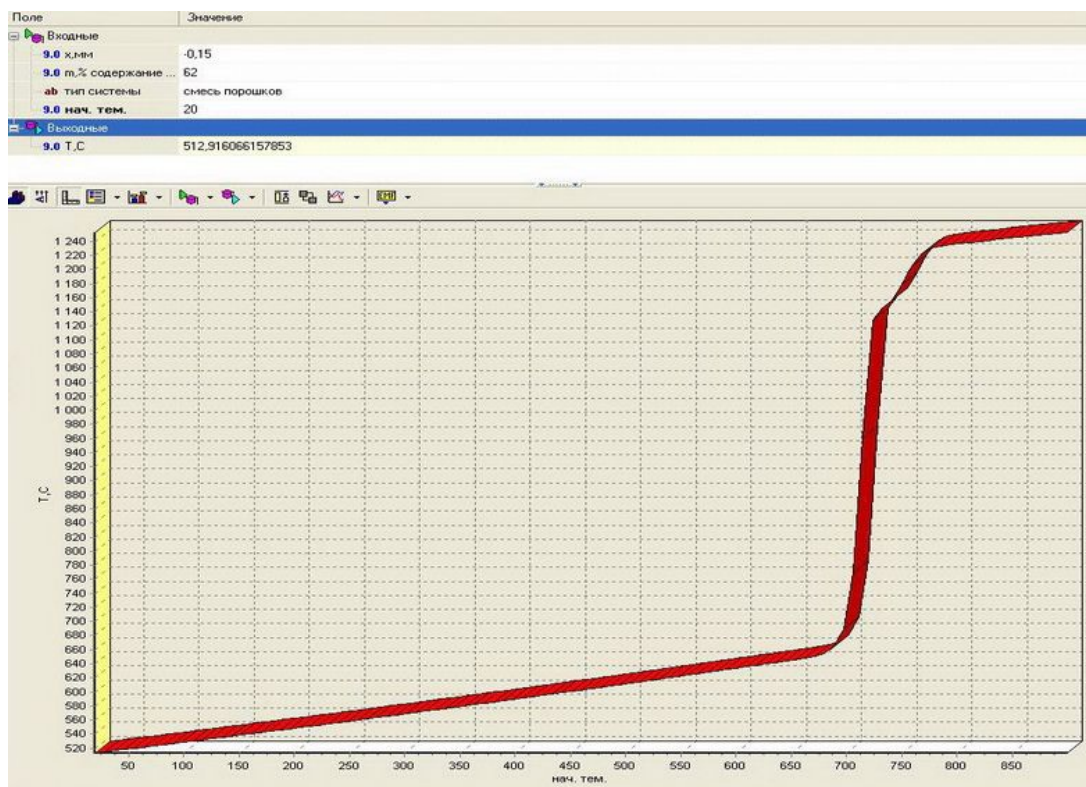


Рис. 5. Зависимость температуры в волне горения (на расстоянии от поверхности горения  $x = -0,15$  мм) от начальной температуры образца (степени предварительного нагрева) для безгазовой КС-смеси порошков Ti, V, Fe при содержании железа – 62 %

Достоинством построенной модели является то, что она позволяет без эксперимента определить температуру в любой точке волны горения и профиль температуры в волне горения в целом для различных составов, содержания железа в системе и начальной температуры системы.

Были получены также другие модели для безгазовых конденсированных систем. Модель 1: скорость горения системы титан-бор (Ti+2V) как трехмерная функция относительной исходной плотности системы, размера частиц титана и начальной температуры. Модель 2: скорость горения системы железо-ванадий (Fe-V) как двумерная функция плотности системы и типа системы. Модель 3: максимальная температура волны горения безгазовых систем как двумерная функция начальной температуры системы и ее типа.

**Заключение.** В статье представлены результаты исследований возможностей создания баз знаний в области физики горения и нанотехнологий с помощью средств Data Mining.

Целью работы было создание комплекса вычислительных моделей горения конденсированных систем и вибрационного горения газов, обобщающих экспериментальные данные, и, на их основе, баз знаний в области горения, а также создание методологических основ для разра-

ботки вычислительных моделей и баз знаний в других областях научных исследований, в частности в области нанотехнологий.

Полученные результаты показывают хорошие перспективы применения средств Data Mining для достижения поставленной цели и решения разнообразных задач моделирования горения конденсированных систем, вибрационного горения, а также свойств продуктов нанотехнологий.

Средства Data Mining позволяют существенно увеличить значение полученных ранее экспериментальных результатов, а также получать новые «экспериментальные» результаты и выявлять новые неизвестные ранее закономерности.

Мы полагаем, что Data Mining будут рассматриваться в будущем как хорошее средство анализа многомерных экспериментальных функций, как средство обобщения и прогнозирования закономерностей процессов горения, связей между переменными задачи.

Полученные вычислительные модели, объединенные в общую аналитическую платформу, можно рассматривать как удобную в использовании базу знаний закономерностей исследуемых объектов, содержащую в себе в компактном виде большое количество разнообразной информации об объекте и обеспечивающую быстрый доступ к ней в удобном графическом виде. Она способна легко адаптироваться к новым экспериментальным данным (дообучаться). Часть ее в виде ИНС-моделей можно представить как своеобразный специализированный компактный инженерный калькулятор, позволяющий мгновенно производить требуемые вычисления характеристик горения конденсированных систем и вибрационного горения, состава топлива, способов влияния на режим горения.

Стоит также отметить интересную роль, которую могут иметь полученные модели в учебном процессе. Обобщая закономерности экспериментальных данных, данные модели позволяют в компактном виде наглядно продемонстрировать особенности механизма горения конденсированных систем, механизма возникновения и прекращения вибрационного горения.

Помимо чисто научных, прикладных и учебных задач, полученные модели позволят облегчить решение и практически важных задач. Современные электронные технологии позволяют создавать копии обученных искусственных нейронных сетей в виде микропроцессоров (чипов), мгновенно обрабатывающих поступающие входные сигналы и выдающих управляющие решения (выходные сигналы). Поэтому искусственные нейронные сети могут рассматриваться как очень перспективное средство для решения задач управления и создания систем управления процессом горения.

Таким образом, в настоящее время Data Mining можно рассматривать как мощное средство, поддерживающее тесное взаимодействие эксперимента, теории, компьютерного моделирования и инженерных расчетов.

### Литература

1. Дюк В., Самойленко А. Data mining: учебный курс. – СПб: Питер, 2001. – 368 с.
2. Круглов В.В., Борисов В.В. Искусственные нейронные сети. – М.: Горячая линия – Телеком, 2002. – 382 с.
3. Neural Networks for Instrumentation, Measurement and Related Industrial Applications. Proceedings of the NATO Advanced Study Institute on Neural Networks for Instrumentation, Measurement, and Related Industrial Applications (9-20 October 2001, Crema, Italy) / ed. by Sergey Ablameyko, Liviu Goras, Marco Gori and Vincenzo Piuri, IOS Press, Series 3: Computer and Systems Sciences, 2003, Vol. 185, Amsterdam.
4. Аналитическая платформа Deductor: [www.basegroup.ru](http://www.basegroup.ru).
5. Abrukov V.S. Artificial Neural Networks and Inverse Problems of Optical Diagnostics / V.S. Abrukov, D.A. Troeshestova, R.A. Pavlov, P.V. Ivanov // Proceedings of the 6th International Conference on Intelligent System Design and Applications. Jinan. China. 2006, October 16-18. – P. 850-855.
6. Abrukov V.S., Schetinin V.G., Troeshestova D.A., Deltsov P.V. Perspectives for Decision of Some Hydrodynamical Problems by Neural Networks Models and Methods. / V.S. Abrukov, V.G. Schetinin, D.A. Troeshestova, P.V. Deltsov // Proceedings of the International Summer Scientific School «High Speed Hydrodynamics» / Ed. by G.G. Cherny, M.P. Tulin, A.G. Terentiev, V.V. Serebryakov, Cortana Corporation, USA – Russia. Washington – Cheboksary, 2002. – P. 391-394.
7. Абриков В.С., Николаева Я.Г., Макаров Д.Н., Сергеев А.А., Карлович Е.В. Применение средств интеллектуального анализа данных (Data Mining) для исследования неполно определенных систем // Вестник Чувашского университета. – 2008. – № 2. – С. 233-241.
8. Abrukov V.S. Application of Artificial Neural Networks for Solution of Scientific and Applied Problems for Combustion of Energetic Materials / V.S. Abrukov, D.A. Troeshestova, A.S. Chernov et al // Advancements in Energetic Materials and Chemical Propulsion / Ed. By Kenneth K. Kuo and Juan Dios Rivera, Begell House Inc. of Redding, USA. Connecticut. – 2007. – P. 268-283.
9. Абриков В. С. Методика создания баз знаний в области вибрационного горения средствами Data Mining [Электронный ре-

- курс]: презентация. – Чебоксары, 2009. Режим доступа: [http://library.chuvsu.ru/inform\\_res/tech\\_lib/downloads/ABRUKOV-vibrgorenie.zip](http://library.chuvsu.ru/inform_res/tech_lib/downloads/ABRUKOV-vibrgorenie.zip).
10. Абриков В. С. Методика создания баз знаний в области горения энергонасыщенных конденсированных систем средствами Data Mining [Электронный ресурс]: презентация. – Чебоксары, 2009. Режим доступа: [http://library.chuvsu.ru/inform\\_res/tech\\_lib/downloads/ABRUKOV-gorenieKS.zip](http://library.chuvsu.ru/inform_res/tech_lib/downloads/ABRUKOV-gorenieKS.zip).
  11. Абриков В. С. Применение искусственных нейронных сетей при решении научных и прикладных задач [Электронный ресурс]: Метод. пособие к курсу лекций «Искусственные нейронные сети»: презентация. – Чебоксары, 2009. Режим доступа: [http://library.chuvsu.ru/inform\\_res/tech\\_lib/downloads/Abrukov-Kazan-2005.zip](http://library.chuvsu.ru/inform_res/tech_lib/downloads/Abrukov-Kazan-2005.zip).
  12. Методическое пособие: Методика применения средств Data Mining. Режим доступа: <http://www.chuvsu.ru/2008/proekt.html>.

# ГИБРИДНЫЕ СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ НА ОСНОВЕ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ

*Болотова Л.С., профессор, Московский институт радиотехники, электроники и автоматики (Технический университет),*

*Боев Б.В., зав. лаб. эпидемиологической кибернетики НИИЭМ им Н.Ф. Гамалеи РАМН, Иваненко И.Ю., студент, Московский институт радиотехники, электроники и автоматики (Технический университет), г. Москва*

Как известно, термин «Гибридная система» (ГС) предполагает отказ от взгляда на объект исследования как на простую однородную сущность и принятие понимания её сложности, составности и неоднородности. Термин «гибрид» понимается как система, состоящая из двух или более интегрированных подсистем, каждая из которых имеет различные языки представления знаний, информации и методы решения. Применительно к системам поддержки принятия решений (СППР) это означает:

- естественную интеграцию различных парадигм моделирования;
- сращивание информации различной модальности (символьной, текстовой, картографической, и т.п.);
- интеграцию различных типов моделей представления знаний;
- интеграцию различных методов и моделей решения подзадач;
- «безшовную» интеграцию знаний и традиционной обработки;
- применение различных типов инструментальных программных средств.

В данном случае под СППР понимается система, обладающая интеллектом, сравнимым с интеллектом человека по качеству принимаемых решений. Этому требованию, на наш взгляд, соответствует наиболее общее определение: гибридная интеллектуальная система (ГИС) есть система, в которой для решения задачи используется более одного метода имитации интеллектуальной деятельности человека.

В докладе рассматривается СППР такого типа для противодействия эпидемиям. В её структуре задействованы следующие парадигмы:

- искусственный интеллект (ИИ);
- аналитические модели на основе нелинейных дифференциальных уравнений;
- Data Mining;
- традиционная обработка информации.

В рамках парадигмы ИИ используются:

- экспертные системы (ЭС);
- метод ситуационного анализа и проектирования модели предметной области;
- нечёткие ЭС;
- искусственные нейронные сети.

В рамках парадигмы Data Mining используются методы:

- проектирования Хранилищ Данных;
- классификации на основе дерева решений;
- кластеризации на основе самоорганизующихся карт Кохонена;
- ассоциации.

Программный инструментарий включает следующие инструментальные средства:

- программа работы с матрицами авиатрафика пассажирских перевозок;
- аналитическая платформа Deductor;
- программа прогнозирования эпидемии на основе уравнений Кермака-МакКендрика;
- система нечёткого моделирования;
- инструментальная система «Помощник Эксперта»;
- инструментальная экспертная система «Решатель проблем».

Последние две в списке системы применялись для построения модели предметной области противодействия эпидемиям, главным образом, для проявления социального контекста принимаемых решений.

За основу моделирования процесса распространения эпидемий был взят пассажиропоток воздушного транспорта, как наиболее быстрый для переноса вирусов в планетарном масштабе, а в качестве примера эпидемии рассмотрен грипп. Основными её составляющими являются:

- подсистема сбора и анализа информации о воздушных полетах;
- БД городов мира с необходимыми характеристиками;
- подсистема прогнозирования хода эпидемии гриппа;
- система поддержки принятия решений и выдачи рекомендаций по выбору мер противодействия эпидемии;
- база нечётких знаний (БНЗ), включающая правила оценки эпидемиологической опасности территории (города, района и т.п.).

Структура системы для противодействия эпидемиям гриппа (СПЭГ), а также подсистема сбора и анализа информации о воздушных полетах; БД городов мира с необходимыми характеристиками, включая результаты интеллектуального анализа данных, были подробно рассмотрены в работах [2, 3, 5, 6]. Поэтому в данном докладе мы рассмотрим остальные подсистемы.

Разработано множество моделей разной сложности, описывающих динамику распространения эпидемии. Так, в модели SIR среди населения выделяются три группы людей: те, кто восприимчив к заражению (S); те, кто был заражен и в состоянии распространять болезнь восприимчивым людям (I); те, кто выздоровел от болезни и невосприимчив к последующему перезаражению (R). Смена состояний людей является односторонней, только  $S \rightarrow I \rightarrow R$ . Возможно дополнение модели ещё двумя группами людей, у которых болезнь находится в инкубационном периоде (E) и умерших (F). Существует два фундаментальных параметра модели –  $\alpha$  (ежедневная норма инфекции) и  $\beta$  (норма восстановления). Они являются параметрами, которые управляют тем, насколько быстро люди переходят в группы I и R [1, 4, 7]. В этой модели все расчеты по анализу и прогнозу эпидемического процесса выполняются на основе системы нелинейных дифференциальных уравнений с соответствующими начальными условиями эпидемии. Модель SIR описывается дифференциальными уравнениями Кермака-Мак Кендрика:

$$\frac{dS}{dt} = -\alpha SI;$$

$$\frac{dI}{dt} = \alpha SI - \beta I = (\alpha S - \beta)I.$$

Обозначения:  $\alpha$  – ежедневная норма инфекции;  $\beta$  – норма восстановления. Как уже говорилось, смена состояний людей является односторонней, т.е.  $S \rightarrow I \rightarrow R$ , но возможно дополнение состоянием E (инкубационный период) и F (летальный исход).

Из уравнений видно, что если часть населения в зараженной группе первоначально увеличивается (то есть,  $\frac{dI}{dt} > 0$  при  $t = 0$ ), то эпидемия началась. Переход между эпидемией и неэпидемическим распространением болезни происходит, когда  $\frac{dI}{dt} = 0$ . Одновременно анализ второго уравнения показывает, что этот переход наступает, когда  $S_0 = \frac{\beta}{\alpha}$ . Аналогично, пик эпидемии происходит, когда  $S = \frac{\beta}{\alpha}$ , а показатель изменения зараженной группы прекращает увеличиваться и начинает уменьшаться [1, 4].

Существует еще один способ представления этой модели. В нем сделан упор на практические аспекты распространения. К примеру, вводится контактное число  $\gamma$ , где  $\gamma = \frac{\beta}{\alpha}$  является одним из самых важных параметров, определяющих динамику распространения болезни. У контактного числа есть легко понимаемая интерпретация: среднее

число восприимчивых членов населения, среди которых зараженный распространяет болезнь к тому, как долго этот человек сам находился в зараженной группе. Последует ли эпидемия при определенных начальных условиях, может быть рассмотрено с точки зрения контактного числа. Опытным путем установлено, что переход между эпидемическими и неэпидемическими состояниями происходит, когда численность населения в восприимчивой группе обратно контактному номеру. Можно ввести меру восстановления  $\beta$ , которую можно представить продолжительностью болезни [1, 4].

Основным компонентом системы противодействия эпидемиям является сама модель эпидемии. В зависимости от прогноза может проводиться изоляция зараженных, блокировка путей передачи заболевания, вакцинация медперсонала, вакцинация восприимчивой части населения или вакцинация всего населения.

В настоящее время СППР СПЭГ реализована как прототип ЭС, задачей которой является выработка рекомендаций ЛПР. Система строит свои рассуждения исходя из своей БЗ и основываясь на входных данных (рис. 1).

В случае подозрения на возникновение эпидемии в СППР вводятся данные о городе, числе инфицированных, подверженных заражению, лиц в инкубационном периоде, умерших, а также коэффициенты для расчета прогноза. Система выводит на экран:

- рекомендуемое действие;
- масштаб его выполнения;
- рекомендации по поводу ресурса больниц;
- объяснение выданной рекомендации;
- список городов, которые необходимо уведомить о вероятности эпидемии.

На данном этапе реализации СППР может давать рекомендации следующих основных типов действий, список которых может и будет расширяться:

- профилактика;
- вакцинация;
- карантин.

Возможен масштаб выполнения действий для:

- района;
- города;
- области.



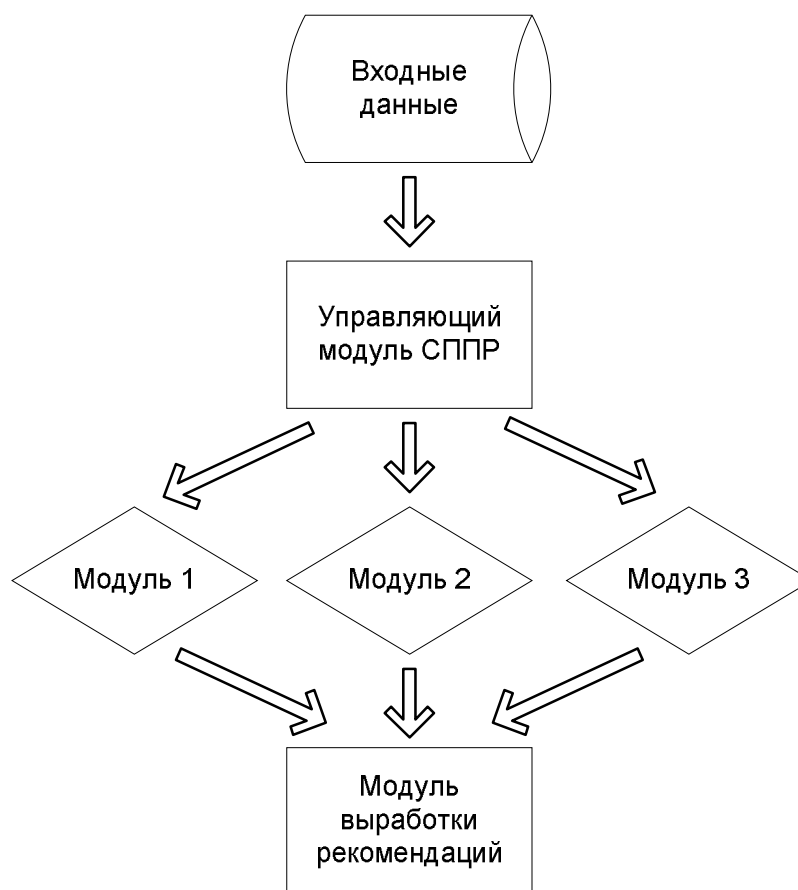


Рис. 1. Общая схема СПДР

Допустимы и другие, более мелкие масштабы при соответствующей подготовке исходных данных.

Основные модули СПДР:

- обработчик входных данных, отвечающий за преобразование входных данных в формат, понятный программе;
- управляющий модуль СПДР, отвечающий за получение данных и дальнейшую их передачу к модулям – обработчикам;
- модули, реализующие задачи прогнозирования и нечеткого вывода.

Модуль анализа авиатрафика ScoreMachine отвечает за анализ БД авиаперелетов. Формат данных в ней определен форматом, выдаваемым подсистемой анализа трафика пассажиров воздушного транспорта. На основе анализа базы находятся города, связанные с рассматриваемым и выдается рекомендация: «Необходимо уведомить следующие города о возможности возникновения эпидемии: (список городов)», рис. 2.

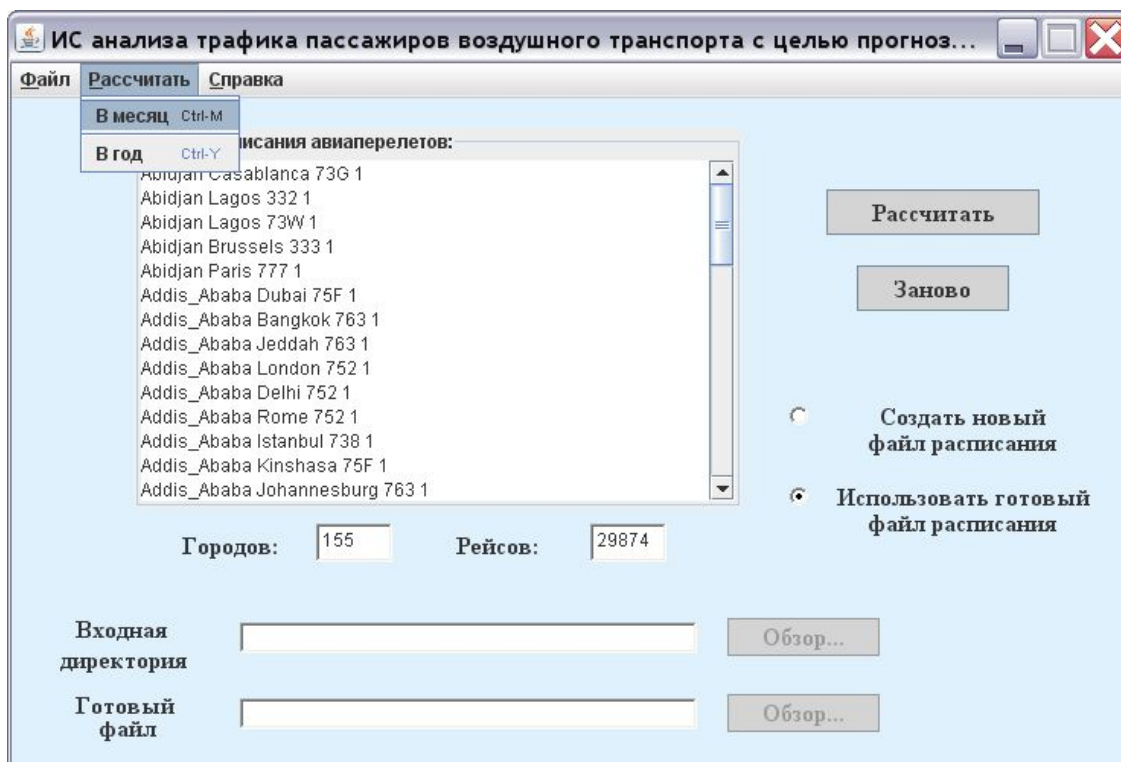


Рис. 2. Окно для расчёта трафика авиапассажиров

Представление знаний в СППР выполнено в виде нечётких правил, а в качестве аппарата их интерпретации, естественно, применяется аппарат нечеткой логики, как наиболее соответствующей характеру знаний экспертов эпидемиологов. При таком представлении экспертам не нужно вводить в БЗ, например, процентные соотношения зараженного населения. Вместо этого достаточно задать этот факт числом, по которому система самостоятельно определит соответствующее нечёткое множество: low (низкое число зараженных); medium (среднее число зараженных); high (высокое число зараженных). Соответствующие этим множествам функции принадлежности (характеристические функции) приведены на рис. 3.

В этом окне пользователю выводятся графики функций для различных состояний. На каждом графике вертикальной чертой обозначается введенное значение. График «solution» показывает решение системы.

Правила БЗ имеют стандартный синтаксис: IF (условие) [OR (условие)] THEN (переменная) IS (значение). Ниже приведен фрагмент правил БЗ:

- RULE 1 : IF incub IS low OR infect IS low THEN solution IS profilactics;
- RULE 2 : IF incub IS low AND infect IS med THEN solution IS vaccination;
- RULE 3 : IF incub IS med AND infect IS low THEN solution IS vaccination;
- RULE 4 : IF infect IS med THEN solution IS vaccination;
- RULE 5 : IF infect IS med AND liable IS low THEN solution IS vaccination;

RULE 6 : IF liable IS med THEN solution IS profilactics;  
 RULE 7 : IF liable IS high THEN solution IS vaccination;  
 RULE 8 : IF liable IS low AND recover is low THEN solution IS vaccination;  
 RULE 9 : IF recover IS low THEN solution IS vaccination;  
 RULE 10 : IF recover IS med and dead is low THEN solution IS vaccination;  
 RULE 11 : IF liable IS low and dead is low THEN solution IS vaccination;  
 RULE 12 : IF dead IS med THEN solution IS quarantine;  
 RULE 13 : IF dead IS high THEN solution IS quarantine.

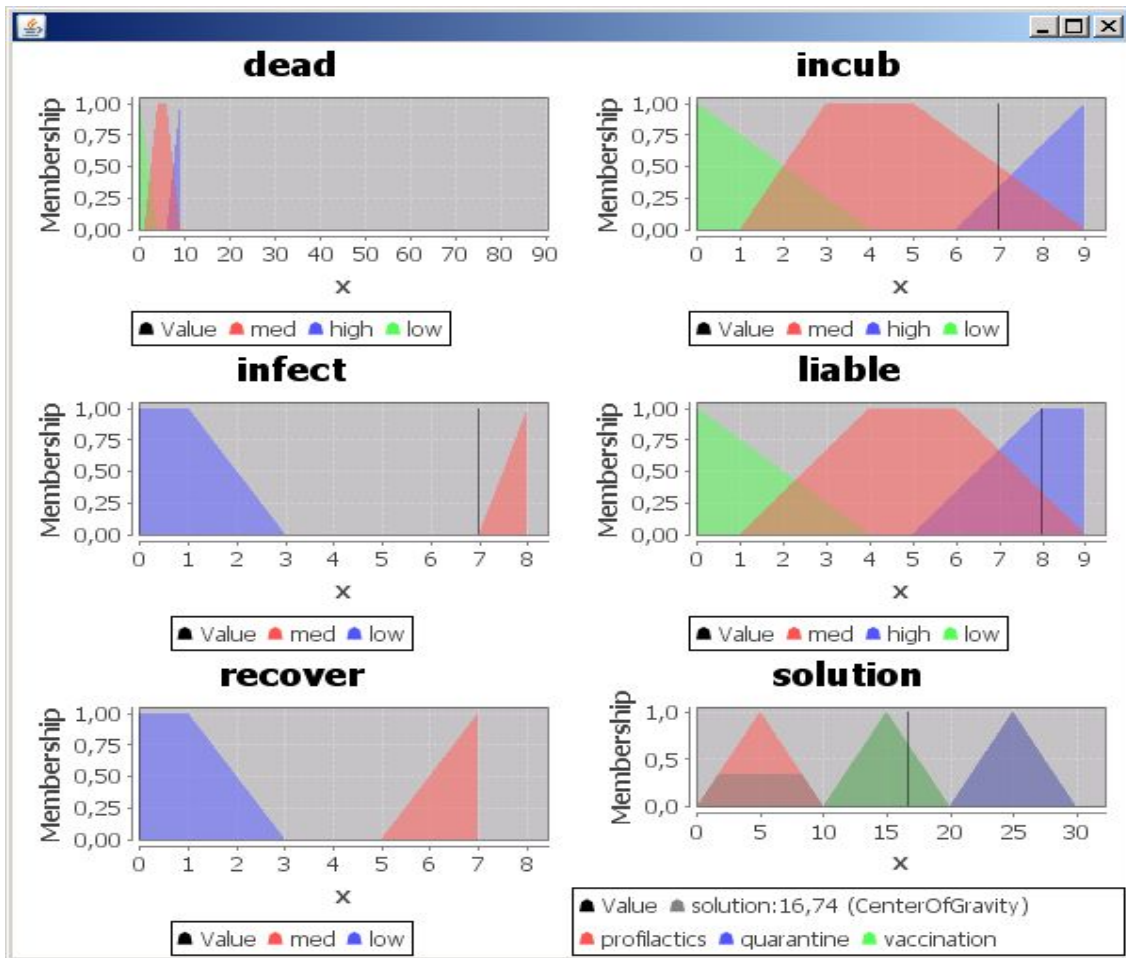


Рис. 3 Окно функций принадлежности нечётких множеств

При работе с программой пользователь заполняет поля «Город», «Число свободных мест в больницах», «Число человек в инкубационном периоде», «Число инфицированных», «Число восприимчивых к заражению», «Число выздоровевших», «Число погибших», «Доля инфекции», «Коэффициент восстановления» (рис. 4).

Управляющий модуль – ядро СППР для противодействия эпидемиям. Он ответственен за обработку входных данных и дальнейшее распределение их по функциональным модулям системы, затем – за обработку возвращенных данных и выдачу их пользователю (рис. 5).

Это основной модуль для получения рекомендаций. На основе данных о численности населения в различных состояниях, БНП выдается рекомендация относительно действия и необходимого масштаба применения этого действия (например, действие – вакцинация, масштаб – город).

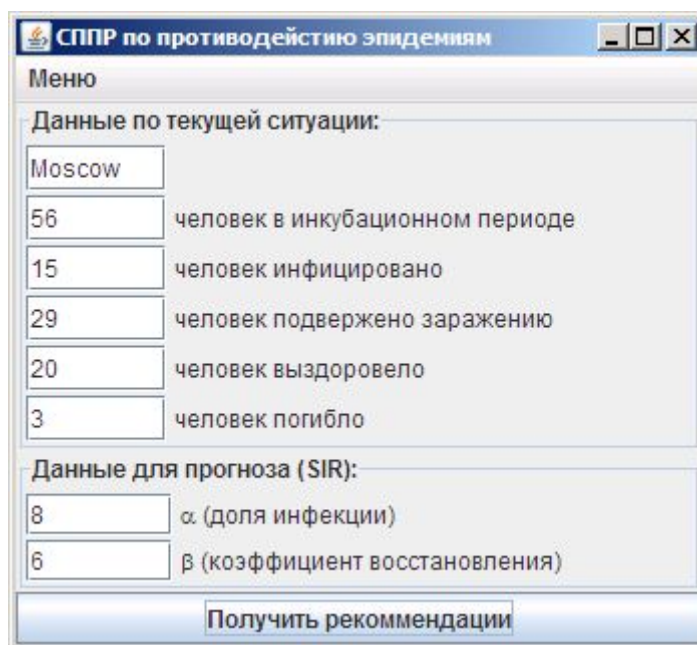


Рис. 4. Окно для представления исходных данных

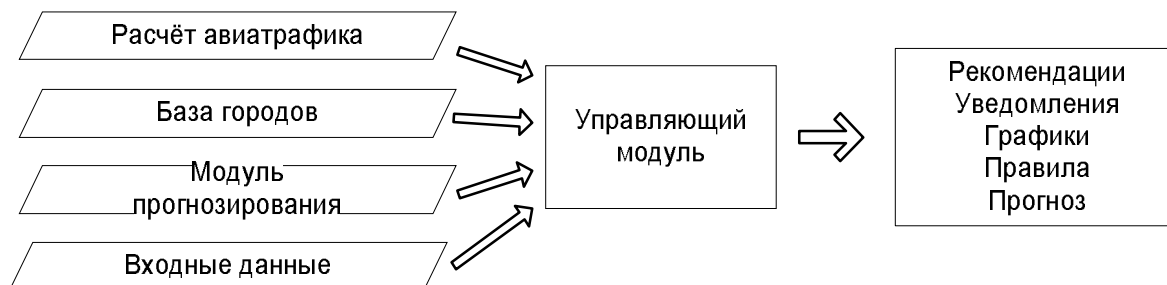


Рис. 5. Структура управляющего модуля

Модуль расчета прогнозов (SIRMachine) на основе данных пользователя (ЛПП) выводит текущее состояние эпидемии, прогноз её дальнейшего развития и необходимый ресурс для размещения больных (рис 6). На рис. 7 показан прогноз, когда можно ожидать окончания эпидемии относительно текущей ситуации – это 60-й день.

Модуль вывода рекомендаций (Recommendation Builder) используется управляющим модулем (Operating module) и собирает результаты работы всех функциональных модулей, а затем выдает их пользователю в виде окна с отображенными рекомендациями (рис. 8).

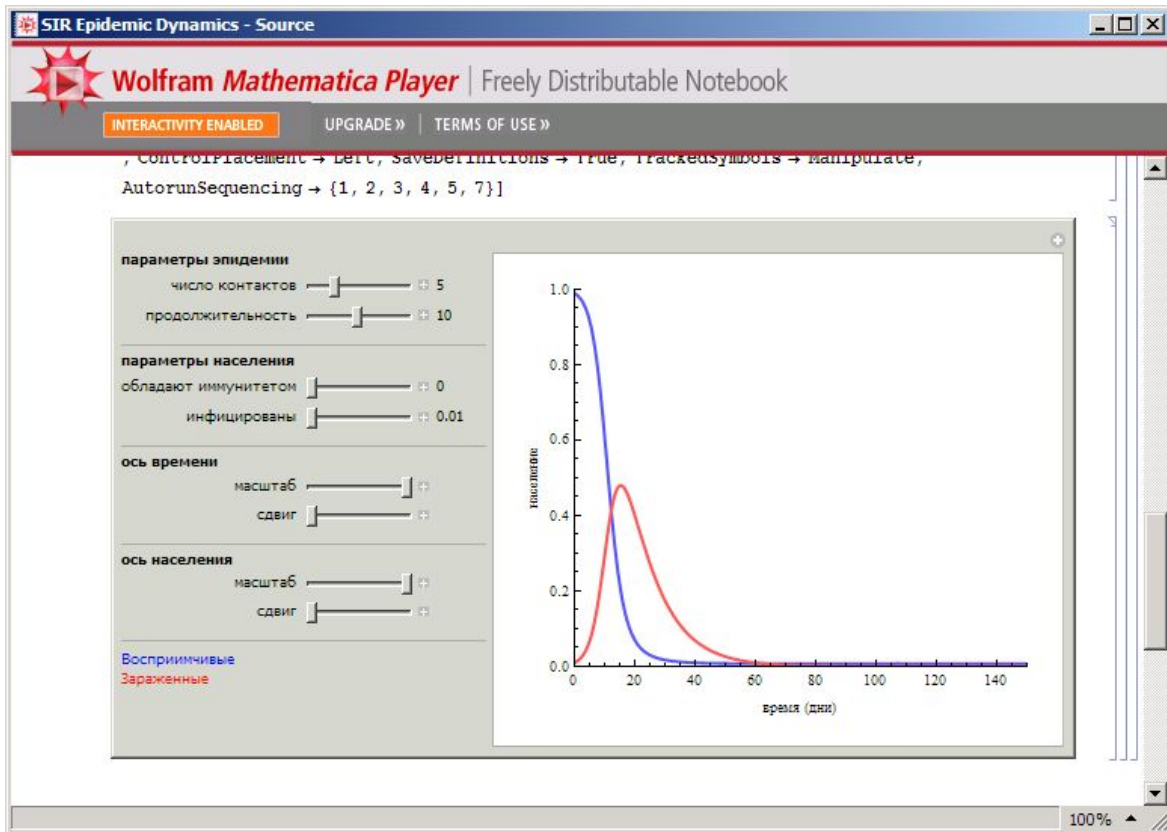


Рис. 6. Текущее состояние эпидемии

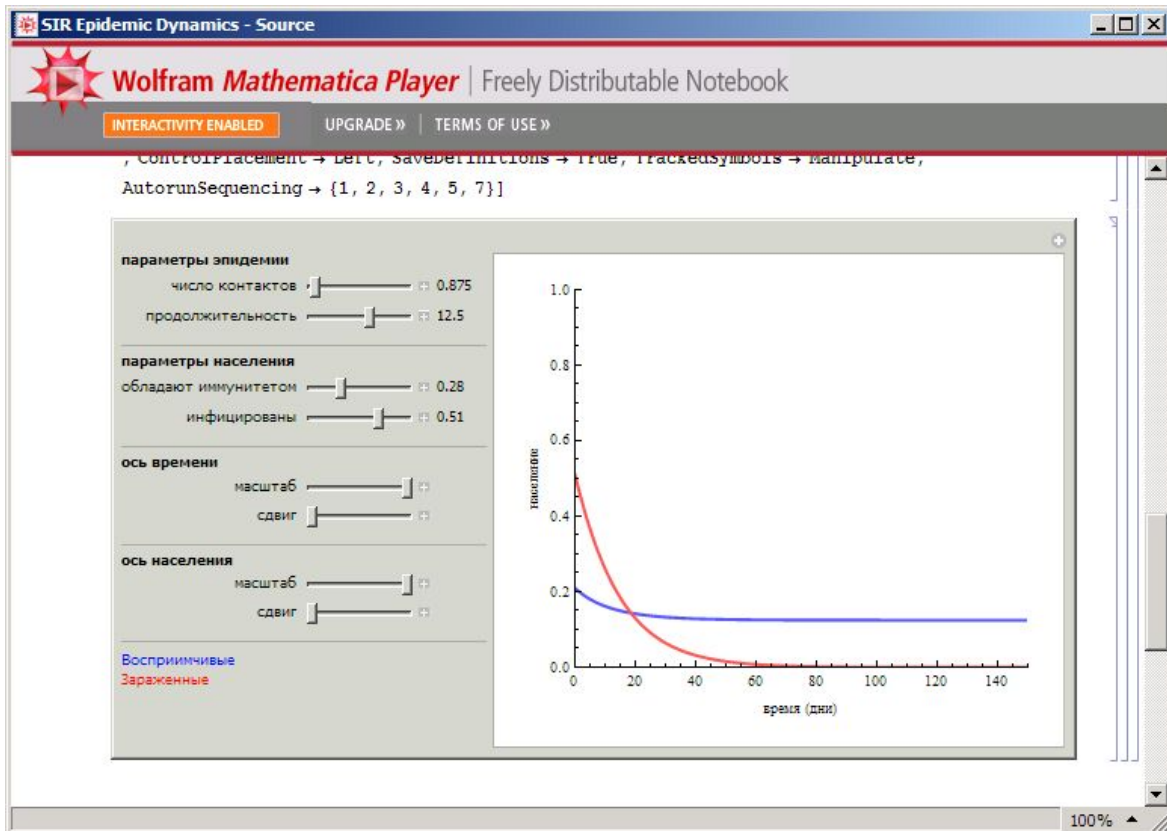


Рис. 7. Прогноз развития эпидемии

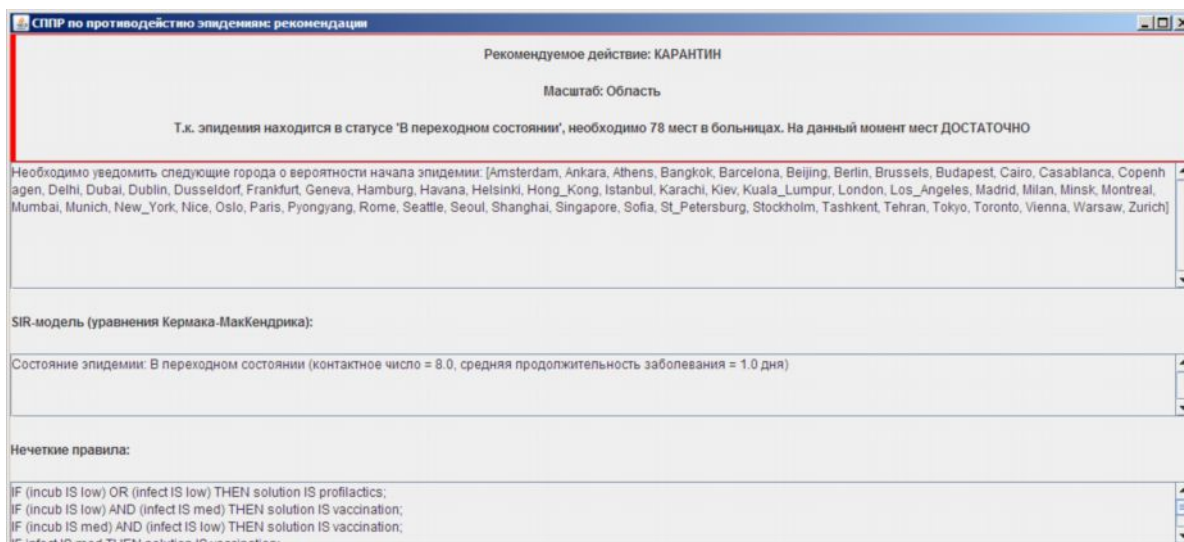


Рис.8. Окно рекомендаций для решения «Карантин»

В окне рекомендаций пользователю выводятся:

- рекомендуемое действие – профилактика, вакцинация или карантин;
- масштаб действия – район, город или область;
- рекомендации относительно ресурса больниц;
- список городов для уведомления о вероятности эпидемии;
- SIR-данные: состояние эпидемии, контактное число и среднюю продолжительность заболевания;
- примененные для вывода решения нечеткие правила.

В заключение подчеркнем следующее.

1. Разработанный прототип прошёл испытания для различных типов решений и показал:

- адекватность принимаемых решений реальным;
- гибкость модели и возможность формирования различных конфигураций на её основе;
- открытость для расширений для: БНП; новых типов решений; новых территориальных уровней;
- возможности расширения предметной области учёта социальной сферы развития эпидемий;
- возможность применения для противодействия другим типам эпидемий, кроме гриппа.

В последнем случае необходима смена модели прогнозирования развития эпидемии.

2. В настоящее время в РФ отсутствует специальная служба для систематического наблюдения и сбора необходимой информации от

различных служб, участие которых необходимо для организованного и эффективного противодействия эпидемиям. Считаем, что на основе данного прототипа может быть создан проект информационного обеспечения такой службы с указанием:

- перечня необходимых БД;
- исследуемых параметров;
- структуры хранилища данных;
- технологии разработки полноценной СППР для поддержки службы противодействия эпидемиям.

### **Литература**

1. Боев Б.В. Современный этап математического моделирования в эпидемиологии инфекционных заболеваний // Сборник научных трудов «Эпидемиологическая кибернетика: модели, информация, эксперименты» – М., 2010. – С. 6 -12.
2. Боев Б.В, Болотова Л.С, Дёмина Н.Н. Интеллектуальный анализ данных в системе противодействия распространению эпидемий гриппа // Бизнес-аналитика. Вопросы теории и практики. Использование аналитической платформы Deductor в деятельности учебных заведений: сборник материалов межвуз. науч.-практ. конф. – Рязань: Лаборатория баз данных, 2010 – С. 134-144.
3. Болотова Л.С., Кузнецов С.Н. Демина Н.Н. Проблемы применения методов интеллектуального анализа данных в системах поддержки принятия решений // Бизнес-аналитика. Вопросы теории и практики. Использование аналитической платформы Deductor в деятельности учебных заведений: сборник материалов межвуз. науч.-практ. конф. – Рязань: Лаборатория баз данных, 2010 – С. 62-66.
4. Иваненко И. Ю. Система поддержки принятия решения для противодействия развитию эпидемий // Сборник материалов 4-й Всероссийской конференции студентов, аспирантов и молодых ученых. М.;, 2010. – Ч. 2. – С. 16-18.
5. Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных к знаниям. – СПб.: Питер, 2009. – 624 с.
6. <http://www.basegroup.ru>.
7. <http://www.gamaleya.ru>.

## **О НЕКОТОРЫХ ТЕХНИЧЕСКИХ ПРИЛОЖЕНИЯХ МЕТОДОВ DATA MINING В ОБЛАСТИ ПРИМЕНЕНИЯ И ЭКСПЛУАТАЦИИ РАКЕТНО-КОСМИЧЕСКОЙ ТЕХНИКИ**

*Галаев С.А., преподаватель, Потюпкин А.Ю., начальник кафедры,  
Военная академия РВСН имени Петра Великого, г. Москва*

На сегодняшний день одним из важнейших направлений развития научно-технического прогресса является космос, что, в свою очередь, влечет за собой совершенствование и усложнение объектов ракетно-космической техники. Любой наверняка знаком с такими понятиями как спутник, ракета, космический аппарат, но мало кто представляет себе, насколько данные объекты сложны. Являясь «вещью в себе» они формируются совокупностью различного рода подсистем, также являющихся до конца непознанными «черными ящиками», то есть обладающих непредсказуемостью своего поведения, что, несомненно, причисляет их к разряду сложных систем. Поэтому они сегодня представляют уже не только теоретический интерес [1].

Сложными являются практически любые системы, обладающие тем, что принято называть скрытыми закономерностями поведения. Как правило, к таким системам относят эргатические системы, в которых очень велико влияние человеческого фактора. Однако и технические системы, сконструированные как «простые», зачастую ведут себя как сложные системы. Помимо «больших» технических систем, в которых число подсистем очень велико, а состав разнороден, сложными являются и относительно небольшие системы, обладающие встроенными средствами ЭВТ с развитым программным обеспечением. Такие системы, функционируя большей частью по известным алгоритмам, и являясь для наблюдателя достаточно «простым» объектом, вдруг начинают вести себя непредсказуемо и быстро становятся «сложными». На практике всякое отступление от заданных алгоритмов действий является нештатной ситуацией и может быть отнесено к проявлениям сложности системы.

Управление сложной технической системой, как и любой другой, предполагает определение её текущего состояния и предсказание последующего. Однако для сложных систем эта задача является затруднительной или, точнее, некорректной. Некорректность может проявляться в неоднозначности решения, а также в неустойчивости результатов, когда малым возмущениям исходных данных могут соответствовать сколь угодно большие изменения результатов. Предсказать поведение такой



системы абсолютно точно невозможно. Проблема заключается в нелинейности поведения системы, в наличии так называемых точек бифуркации. Здесь бифуркация (от лат. *bifurcus* – раздвоение) – приобретение новых качеств динамической системой при малом изменении ее параметров [3]. Рассмотрим данную проблему несколько подробнее.

В обычном состоянии система находится в состоянии гомеостатического равновесия с окружающей средой. Гомеостаз (от греческих *homioios* – подобный и *stasis* – состояние) означает свойство живых организмов поддерживать состояние своей «внутренней среды» постоянным с тем, чтобы находящиеся в ней живые клетки имели возможность функционировать вне зависимости от внешних условий, в которых организм находится. В самом общем понимании гомеостатичность системы означает, что она не только выполняет предписанное ей задание, но одновременно заботится и о сохранении самой себя (по крайней мере, до момента выполнения этого задания). При этом в пространстве параметров внешней среды существует гиперповерхность, отделяющая область их значений, внутри которой возможно существование системы. Эта гиперповерхность носит наименование границы гомеостаза. Переход из области гомеостаза через её границу означает прекращение существования данной системы. В этом случае система попадает в неравновесное состояние (состояние кризиса) или так называемую точку бифуркации, выход из которой может привести как к гибели системы, так и её качественному обновлению (рис. 1). Здесь выход 1 из точки бифуркации ведет к гибели системы, выход 2 – к её обновлению.

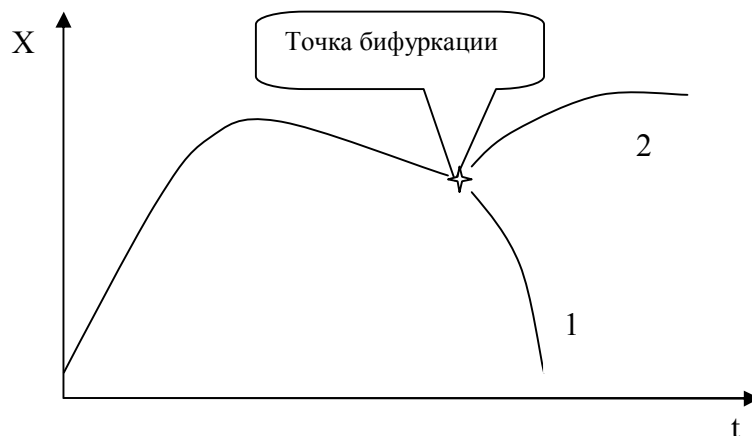


Рис. 1. Кризис системы

Выход системы из кризиса связан с явлением самоорганизации, изучением которого занимается синергетика.

Исследование естественных систем показало, что для реализации процесса самоорганизации требуются следующие условия: наличие не-

устойчивости, наличие флуктуаций (случайный толчок) и наличие возможности выбора состояния системой. В результате система стремится упорядочить своё поведение таким образом, чтобы достичь равновесного состояния по отношению к внешней среде и обеспечить минимальную собственную энтропию.

На рис. 2 показаны различные варианты поведения систем. Для различных систем одни и те же точки могут быть как устойчивыми, так и неустойчивыми. Так для системы S1 точка 2 является минимумом потенциала, соответствует требованию минимального расхода энергии и является точкой устойчивости. Но для системы S2 эта же точка является точкой бифуркации, из которой она может перейти либо в 1, либо в 3 устойчивые состояния. Если рассматривать системы S1, S2 как одну и ту же систему в различные моменты жизни, то тогда ясно, что в процессе жизнедеятельности системы в силу воздействия внешних факторов система не может существовать в прежнем состоянии 2, а вынуждена перейти либо в состояние 1, либо в 3.

Интересно, что термин «бифуркация» [3] происходит от английского «fork» – вилка, так как кривая для S2 имеет форму вилки.

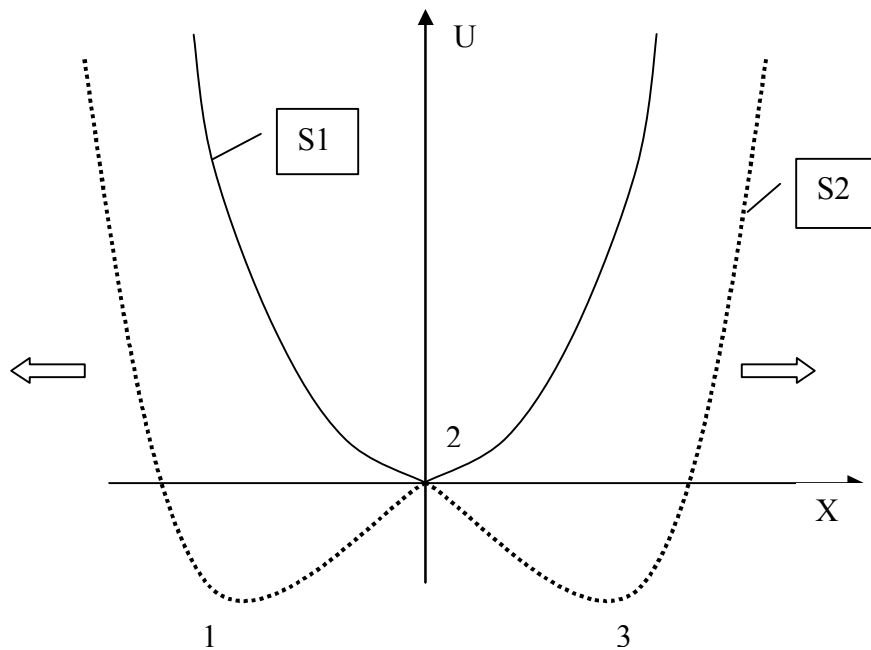


Рис.2. Различные варианты поведения систем

Таким образом, переход системы из одного состояния в другое происходит в результате совместного воздействия стохастических и детерминированных сил (случайность и необходимость), определяющего возникновение новых конфигураций. Для оценки состояния системы требуется определить параметры порядка системы, что предполагает знание модели системы.

Для наблюдателя крайне важным является следующий вопрос: какую же конфигурацию примет система? Например, на рисунке для наблюдателя немаловажно, где же окажется система – в состоянии 1 или в состоянии 3. С точки зрения самой системы состояния равноценны, они обеспечивают и минимум расхода энергии и минимум энтропии, но с точки зрения пользователя – отнюдь нет.

Рассмотрим вариант формализации данной задачи. При этом воспользуемся элементами квантовой механики [3, 4], которые позволяют описать стохастический характер изменения состояний системы. В этом случае состояние системы в момент бифуркации будет описываться линейной суперпозицией её возможных состояний

$$F = \alpha * f_1 + \beta * f_2 + \gamma * f_3 \dots,$$

где  $\{f_1, f_2, f_3, \dots\}$  – множество возможных состояний системы,  $\alpha, \beta, \gamma$  – так называемые амплитуды вероятностей, квадрат модуля которых есть вероятность того, что после проведения измерений система будет находиться в данном состоянии, например  $P(f_1) = |\alpha|^2$ .

При этом возможные состояния являются взаимно ортогональными, а набор возможных альтернатив, отобранный в результате полного измерения, образует систему базисных ортогональных векторов. Для описания подобного рода систем используется Гильбертово пространство, в котором каждый из лучей соответствует определенному квантовому состоянию (рис. 3).

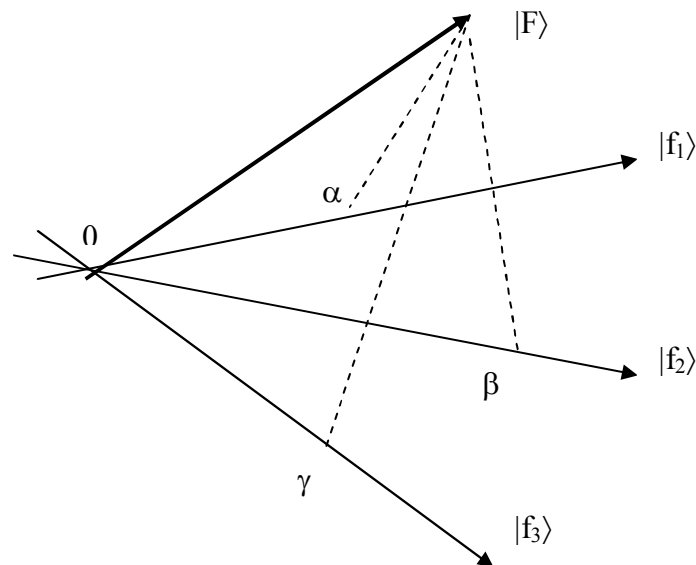


Рис. 3. Описание системы в Гильбертовом пространстве

В обозначениях гильбертова пространства состояние системы будет записываться следующим образом

$$|F\rangle = \alpha * |f_1\rangle + \beta * |f_2\rangle + \gamma * |f_3\rangle \dots$$

В результате измерения состояние системы скачком переходит на одну из осей набора, соответствующего данному измерению. Причем выбор оси происходит случайным образом. Не существует динамического закона, который сказал бы нам, какая из осей будет выбрана. Её выбор случаен, а значения вероятностей определяются квадратом модуля амплитуды вероятности. Задача наблюдателя в случае предсказания заключается в оценке возможных альтернатив и выборе наиболее вероятной (рис. 4).

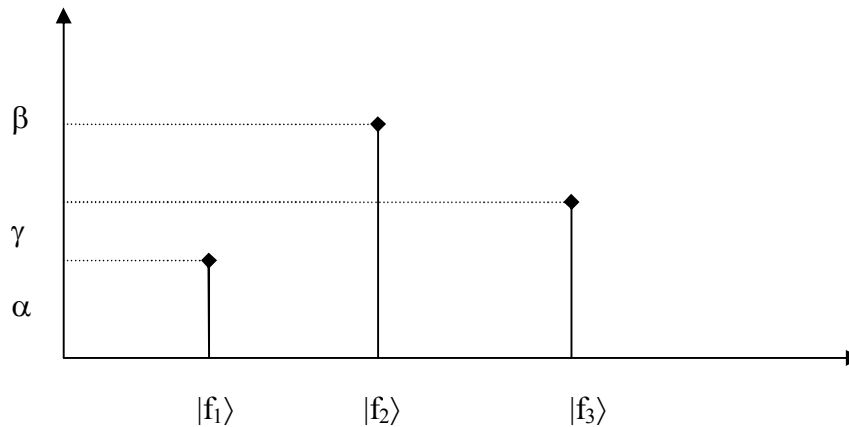


Рис. 4. Альтернативы в Гильбертовом пространстве

Вместе с тем, говорить о полной неопределённости процесса перехода, по-видимому, некорректно. Так, академик Н.Н.Моисеев предполагает, что этот переход обусловлен принципом минимума диссипации, то есть отбор возможных состояний происходит на основе минимума расхода внешней энергии, хотя, как показал И. Пригожин, существуют системы, для которых этот принцип не работает.

Будем исходить из того, что имеем дело со сложными техническими системами, которые изначально были задуманы по своей природе и создавались как простые. Сложность их поведения обусловлена неучетом ряда факторов при проектировании, ошибками в процессе управления и эксплуатации, влиянием внешних факторов, в силу чего они из штатных состояний переходят в нештатные и, тем самым, из простых систем превращаются в сложные. Примем, что для таких систем, действующих на макроуровне и подчиняющихся физическим законам (в первую очередь, законам классической физики), принцип минимума диссипации является основным. Тогда, оценка возможных альтернатив поведения системы должна предполагать и оценку состояния внешней среды.

Исходя из изложенного выше, прогноз состояния сложной системы должен включать в себя две составляющие: прогноз сохранения гомеостаза и прогноз выхода из точки бифуркации, и должен производиться на основании анализа состояния, как самой системы, так и внешней среды.

При каких условиях решается эта задача, существуют ли на сегодняшний день средства её решения? Представляется, что да, существуют.

Для её решения необходима система сбора и обработки информации (система оценки состояния), которая бы обеспечила наблюдаемость возможных состояний объединённой системы «объект – внешняя среда» [3, 4]. При этом принципиальным является выполнение требования *информационной избыточности* системы. Данное требование обусловлено существенно некорректным характером решаемой задачи. Известно, что решение некорректных задач производится их регуляризацией путем целенаправленного привлечения дополнительной информации для формирования т.н. насыщенного носителя информации об объекте. Следовательно, система оценивания состояния сложной системы должна обеспечивать *избыточность измерительную* – в пространстве измеряемых параметров, *избыточность алгоритмическую* – в пространстве алгоритмов обработки измерительной информации и *интеллектуальную избыточность* – в пространстве методов анализа измерительной информации.

Рассмотрим возможные подходы к реализации указанных требований. Представляется, что требование *избыточности измерительной* может быть реализовано за счет технологий «умная пыль», требование *избыточности алгоритмической* – за счет технологий *измерительного программирования*, *избыточности интеллектуальной* – за счет технологий *интеллектуального анализа данных*.

«Умная пыль» – термин, используемый для описания сети из малых беспроводных микроэлектромеханических систем и дополнительных устройств, которые могут взаимодействовать между собой и получать данные о состоянии внешней среды (например, температуре, освещённости, давлении...). Предполагается, что базовые элементы «умной пыли» – моты (англ. moth – мошка), в итоге будут размером с песчинки или даже частицы пыли. Каждый мот имеет собственные сенсоры, вычислительный узел, коммуникацию и питание. Устройства умной пыли будут основаны на низковольтной и глубоко низковольтной наноэлектронике и включать микроисточники энергии вместе с твердотельными импульсными суперконденсаторами (наноионные суперконденсаторы). Недавние разработки в области нанорадио могут использоваться как технологическая база для воплощения умной пыли на практике (рис. 5).

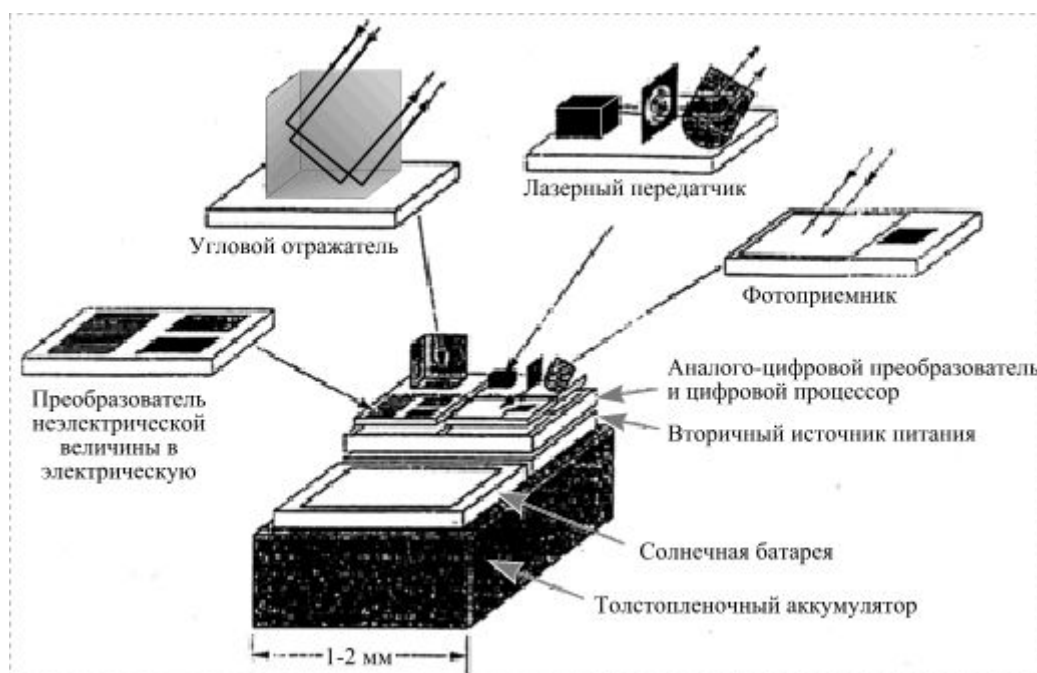


Рис. 5. Элемент «умной пыли».

Выполненные на отдельных кристаллах кремния первичный преобразователь неэлектрической величины в электрическую, аналого-цифровой преобразователь, цифровой процессор совместно с запоминающим устройством, вторичный источник питания и солнечная батарея смонтированы на аккумуляторе, изготовленном в виде бескорпусной толсто пленочной структуры. По вычислительным возможностям каждый датчик, как считают его разработчики, в перспективе будет соответствовать процессору «Intel 8086». Устройства внешней связи представлены блоками фотоприемника, лазерного передатчика и уголкового отражателя. В настоящее время обсуждаются различные методы, которые могут свести размеры мотов умной пыли в сетях датчиков от более миллиметра к микрометрам.

Группируясь вместе, моты автоматически создают очень гибкие сети с малым потреблением питания.

Такая сеть представляет собой «слой», позволяющий организовать обмен данными между двумя другими «слоями» – датчиками, с одной стороны, и информационной системой, в рамках которой они функционируют, с другой. Каждая «пылинка» представляет собой узел беспроводной сети обмена данными с ультранизким энергопотреблением. Передача данных осуществляется от узла к узлу, аналогично тому, как происходит передача пакетов в сети интернет – за исключением того, что в системе «умной пыли» применяется вместо TCP/IP иной протокол передачи данных вкупе со специальным программным обеспечением. Еще одно отличие в том, что разработана технология, позволяющая держать устройства в выключенном состоянии большую часть времени. Новая

технология позволила добиться впечатляющего результата – отдельная «пылинка» на батарейках АА без их замены на сегодняшний день может проработать три года. Представляется, что применение датчиков «умной пыли» позволит наряду с информацией о внутреннем состоянии наблюдаемого объекта получать и информацию о состоянии близлежащей внешней среды. Для этого целесообразно создать своего рода «искусственную кожу» объекта, включающую сеть датчиков тех параметров внешней среды, информация о которых позволит высказать гипотезу о варианте поведения системы при условии реализации принципа минимума расхода внешней энергии.

Следующее требование *алгоритмической избыточности* может быть удовлетворено за счет применения технологий *измерительного программирования*, разработанных на их основе сред измерительного программирования, например LabVIEW [3, 5]. Среды измерительного программирования являются логическим продолжением функционально-модульного принципа построения измерительных систем, и могут быть определены как программно-модульные, где отдельные измерительные функции выполняются не техническими устройствами, а программными модулями (рис. 7) [5].

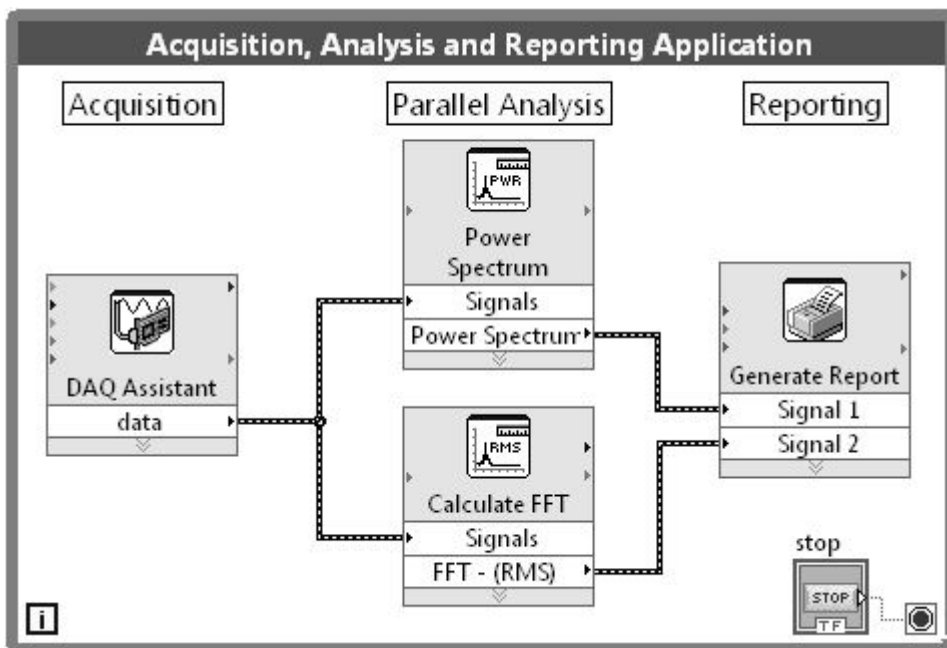


Рис. 7. Пример создания измерительной системы в среде измерительного программирования

Такая среда в рамках системы оценивания позволяет решать следующие задачи:

- организация первичной измерительной системы без внесения технических изменений в объект контроля с целью получения

исходной измерительной и сигнальной информации о его состоянии и состоянии внешней среды;

- формирование информационно-адресной системы регистрации данных с целью их дополнительной обработки методами интеллектуального анализа;
- формирование базы знаний и данных алгоритмов обработки измерительной и сигнальной информации с целью выбора и реализации гибких технологий обработки измерений;
- создание измерительной системы «под задачу»;
- формирование систем обработки с элементами искусственного интеллекта.

Перечисленные возможности сред измерительного программирования позволят оперативно решать задачи обработки избыточного массива измеряемых параметров, полученных в результате измерений датчиками «умной пыли».

Наиболее существенным представляется требование *интеллектуальной избыточности*, которое предполагается выполнить за счет технологий *интеллектуального анализа данных* или технологий Data Mining [2]. Термин *Data Mining* часто переводится как добыча *данных*, извлечение информации, раскопка *данных*, интеллектуальный анализ *данных*, средства поиска *закономерностей*, извлечение знаний, анализ *шаблонов*, «извлечение зерен знаний из гор *данных*», раскопка знаний в базах *данных*, информационная проходка *данных*, «промывание» *данных*. Эти технологии позволяют работать с системами, представляемыми в виде «черного ящика», обрабатывать данные различной природы, причем как количественные, так и качественные, и получать нетривиальные результаты. Для их реализации разработаны и активно используются аналитические платформы – специализированные программные комплексы, в том числе и отечественного производства, которые реализуют алгоритмы очистки, трансформации данных, Data Mining.

К числу задач анализа состояния сложных технических систем [2], решаемых с помощью методов Data Mining, можно отнести следующие:

1. Задача классификации – определение класса объекта по его характеристикам – как задача определения одного из штатных технических состояний системы.
2. Задача регрессии – определение по известным характеристикам объекта некоторых его параметров – как задача факторного анализа.
3. Задача поиска ассоциативных правил – нахождение частых зависимостей (ассоциаций) между объектами или событиями –



как задача определения неявных зависимостей между параметрами системы, задача секвенционального анализа.

4. Задача кластеризации – поиск независимых групп (кластеров) и их характеристик во всем множестве анализируемых данных – как задача выделения и распознавания новых (нештатных) состояний, то есть задача идентификации.
5. С помощью результатов решения всех вышеперечисленных задач можно осуществить и прогноз состояния системы.

Сценарий обработки полученной об объекте измерительной информации позволяет реализовать множество алгоритмов анализа одновременно, получить комплексное заключение по результатам анализа технического состояния и, тем самым, реализовать требование интеллектуальной избыточности на практике.

Анализ ряда успешных применений Data Mining в технических приложениях позволяет утверждать, что методы Data Mining вполне применимы как средство анализа состояния «простой» системы, то есть как средство прогноза сохранения гомеостаза системы, и незаменимы при анализе «сложной» системы как средство прогноза выхода из точки бифуркации.

В целом можно констатировать, что анализ сложной технической системы должен проводиться как анализ системы «объект – внешняя среда». Такой анализ представляет собой крайне сложную задачу и может быть реализован только на принципах информационной избыточности с привлечением разнородных источников информации, разных алгоритмов обработки и анализа, а вся система в целом должна представлять собой интеллектуальную систему. В силу этого проблема анализа сложной системы может быть разрешена путем создания системы анализа с логикой, близкой к человеческому мышлению, способной к восприятию и обработке разнородных источников информации, как синтаксической, так и семантической. Такая система должна быть устойчивой по отношению к влиянию неопределенных факторов для обеспечения корректности анализа, что достигается практической реализацией принципа *информационной избыточности* с применением технологий «умной пыли», измерительного программирования и интеллектуального анализа – Data Mining.

### **Литература**

1. Моисеев Н.Н. Человек и ноосфера. – М.: Молодая гвардия, 1990.
2. Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных к знаниям. – СПб.: ПИТЕР, 2009.

3. Потюпкин А.Ю. Научно-методические основы решения задач анализа состояния объектов ракетно-космической техники в условиях неопределенности – М.: ВА РВСН, 2003.
4. Чечкин А.В. Математическая информатика. – М.: Наука, 1991.
5. [http:// www.labview.ru](http://www.labview.ru)

## РЕШЕНИЕ ЗАДАЧ ОЦЕНКИ КОМПЕТЕНТНОСТИ СТУДЕНТА С ИСПОЛЬЗОВАНИЕМ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ

*Махныткина О.В., преподаватель,  
Алтайский государственный университет, г. Барнаул*

С 2011 года подготовка студентов осуществляется по новым федеральным государственным образовательным стандартам высшего профессионального образования (ФГОС ВПО). Основная особенность новых ФГОСов состоит в том, что они разрабатываются на основе компетентностного подхода. Одним из оптимальных путей формирования компетентности является создание соответствующей модели и встраивание ее в учебный процесс образовательных учреждений системы образования.

Учитывая сложность и неформализованность задач, связанных с оценкой компетентности студента, большое количество первичных факторов, которые необходимо учитывать, можно сделать вывод о необходимости использования специальных методов обработки информации.

Комплексная оценка компетентности выпускника требует разработки соответствующих контрольно-оценочных средств и технологий, методов накопления результатов и аналитических оценок состояния на каждом этапе. При моделировании многокомпонентных, сложных систем и разработке методов структурирования информации традиционно использовали иерархический подход [1] как методологический прием расчленения формально описанной системы на уровни.

При построении моделей первым этапом является системный анализ, позволяющий, по возможности, в максимальной степени структурировать задачу, перевести ее в разряд формализованных. В рамках этого подхода компетентностную модель выпускника можно представить в виде многоуровневой иерархической структуры, а сама задача построения комплексной оценки представляется в виде иерархического дерева подзадач, представленного на рисунке 1. На нижнем уровне оценивается множество частных компетенций, которые не подлежат дроблению на более мелкие и являются основой компетентностной модели.

Модель решения задачи представляет собой ориентированный граф без циклов и петель, узлами которого являются интеллектуальные компоненты  $\{y_{i,j,k,\dots}\}$  с входными и выходными параметрами, а дуги отражают факт непосредственной функциональной зависимости парамет-

ров (узлов) друг от друга. Каждый узел сети может иметь любое конечное количество входящих и исходящих дуг.

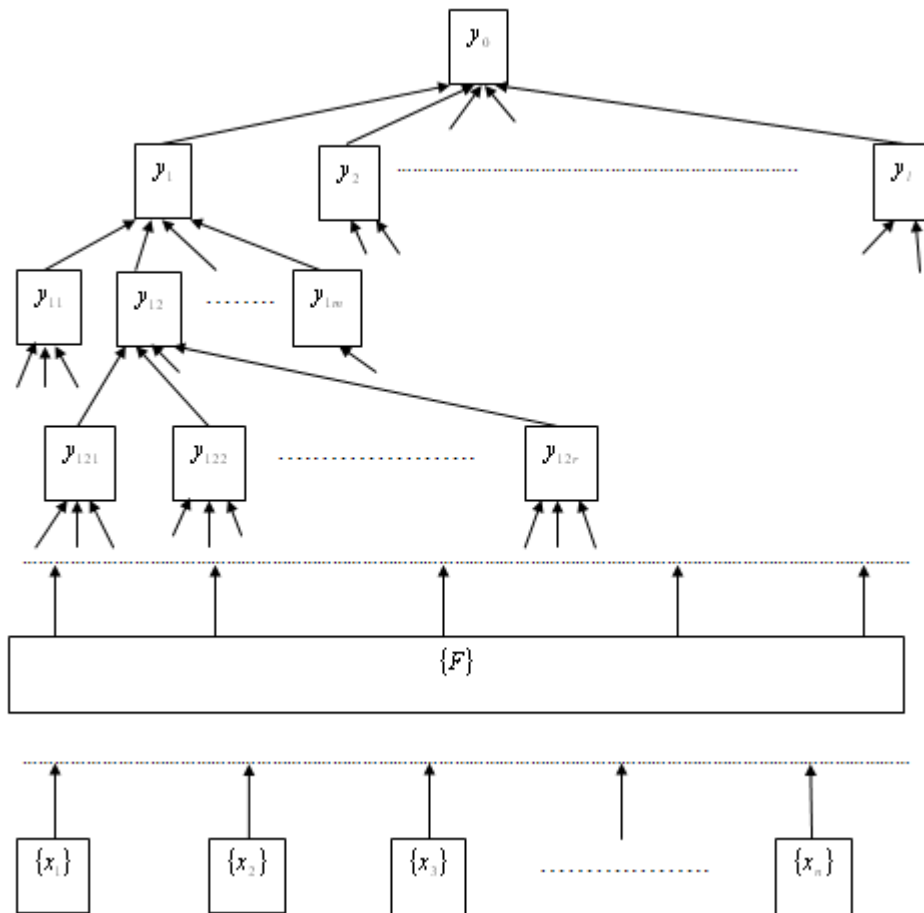


Рис. 1. Представление комплексной задачи оценки компетентности

Пройдя через все уровни графа, выделенные в процессе декомпозиции, на заключительном этапе рассчитывается значение комплексной оценки  $y_0$  при известном векторе частных показателей на основе значений вершин предшествующего уровня. Т.е. в общем виде задача комплексного оценивания компетентности студента ВУЗа может быть представлена в следующем виде.

Обозначим:  $N = \{1, 2, \dots, n\}$  – множество частных критериев, оценки  $x_i \in X_i$  по которым принимают значения из множеств  $X_i$ ,  $i \in N$ ,  $y_0 \in Y_0$  – комплексная оценка, которая вычисляется в соответствии с процедурой агрегирования  $F(\cdot): X' \rightarrow Y_0$  или  $y_0 = F^k(F^{k-1}(\dots F^1(x)))$ , где  $x = (x_1, x_2, \dots, x_n) \in X' = \prod_{i \in N} X_i$ ,  $F_i^k$  – оператор преобразования информации в  $k$ -ом узле.

В процессе прохождения через модель информация сжимается и, в конечном счете, сворачивается в единственный показатель – комплексную оценку. Однако все значения промежуточных вершин являются дос-

тупными. Такая схема агрегирования характерна для социально-экономических объектов.

Решение задачи комплексной оценки уровня сформированности компетентности составляет последовательный интеллектуальный анализ данных.

Оценка уровня сформированности компетенций нижнего уровня производится на основании множество показателей  $x = (x_1, x_2, \dots, x_n)$  (тематической области компетенции), в качестве которых могут выступать результаты устных опросов (собеседование, коллоквиум, зачет, экзамен), письменных работ (тесты, контрольные работы, эссе, рефераты, курсовые работы, научно-учебные отчеты по практике, отчеты по научно-исследовательской работе студентов), анкетирования, компьютерного тестирования, опросов экспертов (преподавателей) и т.д. Социально-личностные характеристики также обусловлены направлением подготовки и могут быть оценены, как по результатам освоения некоторых дисциплин, так при помощи психологических тестов на темперамент, уровень ответственности, конфликтности и др.

В зависимости от характера показателя оценки он может принимать различные допустимые значения и измеряться в разных шкалах (таблица 1).

Таблица 1

Типы квалиметрических шкал для оценки компетентности студентов

Тип шкалы	Показатели оценки	Математические операции с объектами
Номинальная	Результаты тестов на определение социально-личностных характеристик студента и т.д.	Накопление частот (для последующей статистической обработки)
Порядковая	Тесты на определение степени выраженности какого-либо признака, экзаменационные оценки, рейтинги и т.д.	Накопление частот, добавление постоянной, умножение на постоянную
Количественная	Результаты психофизических тестов и т.д.	Любые арифметические действия

Следующим этапом анализа является очистка и предобработка данных, которая заключается в выявлении аномалий и дубликатов, обработке пропусков и нормировании.

Аномальные значения требуют большого внимания при подготовке данных к анализу. Большинство таких значений является ошибкой ввода. Часть аномальных значений вызвана исключительными обстоятельствами, и не отражают реальную ситуацию в исследуемом процессе. С другой стороны, исследование аномального поведения данных позволит прогнозировать условия, вызывающие аномальные события и их послед-

ствия, исследовать реакцию на аномальные изменения условий. Дубликаты также весьма распространенные явления в данных. Дубликатами являются просто одинаковые данные (записи). Они могут дублировать информацию об одном и том же событии, а могут содержать идентичную информацию о двух различных, но похожих событиях. В первом случае дубликаты должны быть просто удалены, а во втором случае требуют более тонкой обработки.

Переменные, используемые для оценки компетентности, могут быть измерены в разных единицах измерения (например, оценка за экзамен и результат тестирования). Чтобы избежать «доминирования» переменных с большим масштабом измерения, предлагается провести предварительную нормировку исходных переменных. Для вещественных переменных могут быть использованы следующие методы нормировки:

$$z_1 = \frac{x - \bar{x}}{\sigma}, z_2 = \frac{x}{x}, z_3 = \frac{x}{x_{\max}}, z_4 = \frac{x - \bar{x}}{x_{\max} - x_{\min}},$$

где  $z_1, z_2, z_3, z_4$  – нормированные значения переменной  $x$ ,  $\bar{x}$  – среднее значение переменной,  $x_{\max}$  – максимальное значение переменной  $x$ ,  $x_{\min}$  – минимальное значение переменной  $x$ ,  $\sigma$  – среднеквадратическое отклонение переменной  $x$ .

После построения модели структуры решения задачи и предобработки производится выбор методов решения в узлах графа связей задач. Основными видами организации функциональной зависимости между параметрами в данной работе являются:

- аналитические зависимости;
- экспертные системы на основе продукционных правил;
- нейросетевые системы.

Выбор метода решения для каждой задачи обуславливается сложностью решаемой проблемы. Автором были выделены характеристики задач, по которым определяется их сложность.

1. Размерность задачи. Под размерностью задачи будем понимать количество рассматриваемых переменных. Размерность задачи определяется тематической областью формирования компетенции. Выделяют задачи малого, среднего и большого размера.

2. Динамичность изменения проблемной области. Выраженная динамичность социально-экономических процессов проявляется в непрерывном изменении их параметров, а то и структуры системы, в которой протекают эти процессы. Часть подзадач оценки компетентности являются динамичными задачами, что определяется следующими факторами:

- компетентностная модель строится на основании требований работодателя, чьи требования к выпускникам постоянно меняются;
- появлением новых учебных дисциплин и дополнительных курсов;
- развитием существующих форм обучения и появлением новых форм обучения.

3. Наличие или отсутствие формальной методики решения задачи, т.е. часть подзадач оценки компетентности может быть решена с помощью существующих, апробированных на практике методик, для решения других необходимо разработка методов решения.

4. Характер входных показателей, в том числе наличие или отсутствие обучающих примеров.

Рассмотрим возможности применения рассматриваемых методов представления знаний в соответствии выделенными критериями сложности задачи.

**Аналитические формулы** можно использовать для решения задач малого и среднего размера, как правило, детерминированных. Построение аналитических зависимостей затруднено для большого числа количественных показателей, а также качественных показателей. Аналитические зависимости имеют ряд преимуществ по сравнению с другими видами моделей. Во-первых, развитые математические методы позволяют преобразовывать выражения для получения наиболее удобного вида модели, характеризующей поведение исследуемой системы. Таким образом, можно провести анализ в общем виде, определить оптимальные соотношения параметров для эффективного управления. Во-вторых, при подстановке численных значений параметров можно контролировать адекватность модели. При этом, однако, для эффективного применения аналитические модели не должны быть слишком сложными.

**Продукционная модель**, или модель, основанная на правилах, позволяет представить знания в виде предложений типа: Если (Условие), то (Следствие) [2]:

$$(Условие\ 1) \ \& \ (Условие\ 2) \dots (Условие\ T) \rightarrow (Следствие)$$

Продукционные экспертные системы можно использовать для разрешимых и трудно разрешимых задач малого размера. В случае рассмотрения новых показателей требуется пересмотр правил. Продукционная модель, основанная на правилах, позволяет использовать как количественные, так и качественные показатели. Продукционные экспертные системы привлекают наглядностью, высокой модульностью, легкостью

внесения дополнений и изменений и простотой механизма логического вывода [3].

Правила продукции разрабатываются с помощью экспертов в данной предметной области. Экспертная деятельность в области образования – система действий, выполняемых с привлечением экспертов, для анализа и оценки качества образовательного процесса с целью повышения обоснованности принимаемых решений в условиях частичной неопределенности, противоречий или конфликтов. В качестве экспертов могут выступать студенты, выпускники вуза, потенциальные работодатели и преподаватели. Одним из существенных условий повышения надежности экспертных оценок социальных объектов является научно обоснованный отбор и формирование экспертной группы. Точность групповой оценки экспертов зависит от численности экспертной группы: уменьшение числа экспертов ведет к снижению точности оценок, т. к. оценка каждого эксперта приобретает больший вес, увеличивается роль субъективного фактора. Однако при большом количестве участников экспертизы усложняется обработка их суждений, становится сложным выявить согласованность их мнений [4].

В качестве инструментария для построения продукционных моделей может быть использован метод построения деревьев решений С4.5. Деревья решений являются наиболее распространенным в настоящее время подходом к выявлению и визуализации логических закономерностей в данных. Каждому узлу сопоставлен некоторый признак, а ветвям – либо конкретные значения для качественных признаков, либо области значений для количественных признаков.

**Искусственные нейронные сети** представляют собой средства вероятностного моделирования данных. Нейронные сети – обобщенное название группы математических алгоритмов, обладающих способностью обучаться на примерах, «узнавая» впоследствии черты встреченных образцов и ситуаций [5].

Нейронные сети особенно продуктивны в решении слабоструктурированных проблем, так как обычно эксперт может легко структурировать проблему до уровня «черного ящика». Нейронные сети позволяют решать задачи, для которых нет информации о методах решения, но есть статистический материал. Таким образом, сети можно использовать для решения трудно разрешимых задач. Они могут оперировать с большим числом показателей. Нейронные сети обладают способностью дообучаться, т.е. могут быть пригодны для изменчивых и нестабильных задач.

Искусственные нейронные сети могут быть использованы для решения ряда типов задач оценки компетентности студента ВУЗа. Первый



тип представляет задачу обучения с учителем и определяется следующим образом.

Пусть в качестве исходных данных имеем матрицу  $R$ :

$$R = \begin{pmatrix} x_{1,1} & \dots & x_{1,n} & y_{1,n+1} & \dots & y_{1,n+q} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{l,1} & \dots & x_{l,n} & y_{1,n+1} & \dots & y_{1,n+q} \\ x_{l+1,1} & \dots & x_{l+1,n} & - & \dots & - \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{l+p,1} & \dots & x_{l+p,n} & - & \dots & - \end{pmatrix}.$$

Первые  $l$  элементов матрицы используется для построения преобразования  $F$ , служащего для восстановления значений  $y_{i,j}, i \in [l+1, l+p], j \in [n+1, n+q]$  по значениям  $x_{a,b}, a \in [l+1, l+p], b \in [1, n]$ . Строки с 1 по  $l$  называют задачиком, в котором выделяют обучающую и тестовую выборки. Обработка при помощи нейросети строк с  $l+1$  по  $l+p$  – процесс решения задачи. В матрице  $R$  строки с 1 по  $l$  содержат сведения о значениях параметров студентов, в столбце  $y$  содержатся экспертные оценки их компетентности. На основе обучения по строкам с 1 по  $l$  программа заполняет столбец  $y$  в строках с  $l+1$  по  $l+p$ , давая оценку другим студентам ВУЗа.

Второй тип представляет задачу обучения без учителя. Сеть такого типа рассчитана на самостоятельное обучение, во время обучения сообщать ей правильные ответы необязательно. В процессе обучения на вход сети подаются различные образцы. Сеть улавливает особенности их структуры и разделяет образцы на кластеры, а уже обученная сеть относит каждый вновь поступающий пример к одному из кластеров, руководствуясь некоторым критерием «близости». К такому типу задач можно отнести обработку данных социологических исследований, тестирование студентов с целью определения их профессиональной направленности. Одним из методов обучения без учителя являются самоорганизующиеся карты Кохонена.

Необходимо заметить, что выбор метода решения для каждого узла условен. Более точно можно определить, попробовав использовать все методы. Метод, дающий наименьшую ошибку при тестировании, является оптимальным для данного узла.

Все этапы решения задачи оценки компетентности были реализованы с использованием аналитической платформы Deductor [6].

Предложенная модель при решении задач оценки компетентности студента дает определенные преимущества. Во-первых, для разных подзадач можно использовать разные методы решения. Часть подзадач про-

ще решить при помощи формальных методик, другие – неформализованные, но имеющие логическую прозрачность – при помощи традиционных экспертных систем. Причем в структуру такой системы могут входить блоки, реализующие математические методы, экспертные системы, основанные на правилах продукций и др., а также нейросетевые элементы. Во-вторых, при изменении в механизме решения одной из подзадач не возникает необходимости в корректировке остальных. В-третьих, можно применить различные методы решения к одной и той же задаче и сравнить результаты решения. В-четвертых, агрегирование информации в рамках гибридной модели не приводят к ее потере, все данные вплоть до первичных могут быть восстановлены.

### **Литература**

1. Месарович М., Мако Д., Такахара И. Теория иерархических многоуровневых систем. М.: Мир, 1973. 344 с.
2. Абдикеев Н.М. Проектирование интеллектуальных систем в экономике: Учебник /Под ред. Н.П. Тихомирова. – М.: «Экзамен», 2004. – 528 с.
3. Базы знаний интеллектуальных систем / Т.А. Гаврилова, В.Ф. Хорошевский. – СПб.: Питер, 2001. – 384 с.
4. Берестнева О.Г., Марухина О.В. Компьютерная система принятия решений по результатам экспертного оценивания в задачах оценки качества образования // Образовательные технологии и общество (Educational Technology & Society) Т. 5, Вып. 3. – Казань, 2002. – С.216-230.
5. Андрейчиков А. В. Интеллектуальные информационные системы: учебник для вузов / А. В. Андрейчиков, О. Н. Андрейчикова. – М.: Финансы и статистика, 2004. – 424 с.
6. <http://basegroup.ru>

## **ИСПОЛЬЗОВАНИЕ DEDUCTOR ДЛЯ ОЦЕНКИ ПРОДОЛЖИТЕЛЬНОСТИ ЖИЗНИ В ОТДЕЛЬНОМ РЕГИОНЕ**

*Машинцов Е.А., доцент, Найденов А.А., аспирант,  
Тульский государственный университет, г. Тула*

На сегодняшний день одной из основных проблем, с которыми сталкиваются муниципальные органы управления, это эффективное использование бюджетных средств, которых, как правило, недостаточно для решения всего комплекса проблем. В связи с этим возникает задача оптимального распределения ограниченных ресурсов для достижения максимальной пользы для общества и, как его составляющей, отдельно взятого человека [1].

Для решения данной задачи управления было бы удобно иметь некую величину для измерения среднего благосостояния населения. Существует несколько показателей качества жизни (далее КЖ), все они в той или иной форме включают такие факторы как: ожидаемую продолжительность жизни, здоровье, семейную и общественную жизнь, материальное благополучие, уровень образования, политическая свобода, климат и др. Следует отметить, что многие из этих факторов являются крайне субъективными, их измерение возможно только посредством опросов и зачастую не существует модели, которая бы могла связать изменение показателя и израсходованные средства. Более того, некоторые из факторов (например, климат или политическая ситуация в стране) не подконтрольны местным органам власти. Т.е. широко используемые индексы качества жизни плохо подходят для принятия управленческих решений.

Среди составляющих интегральных показателей КЖ особое место занимает средняя продолжительность жизни. С одной стороны, это неоспоримо важный фактор для каждого человека, с другой стороны, на него в той или иной мере оказывают влияние все остальные факторы, иными словами, он является неким интегральным показателем КЖ. Достоинством данного показателя является простота и объективность его измерения. Все это позволяет использовать среднюю продолжительность жизни в качестве оптимизируемой величины при решении задач управления на муниципальном уровне.

Прежде чем приступить к задаче оптимизации необходимо выделить и определить силу воздействия факторов, которые влияют на среднюю продолжительность жизни. Это представляется довольно сложной задачей из-за большого их числа, однако в случае принятия управленческих решений на местном уровне можно ограничиться только фактора-

ми, которые поддаются изменению, и рассматривать все остальные в качестве фона.

В данной работе мы предлагаем метод оценки изменения средней продолжительности жизни в отдельном регионе при варьировании тех или иных факторов. В частности мы приводим пример изучения влияния загрязнения воздуха на население г. Тулы с использованием пакета Deductor (более подробный анализ экологической ситуации в г. Тула можно найти в [1]).

Для построения модели территория города была разбита на квадраты размером 500×500 метров. В каждом дискрете была измерена концентрация загрязняющих веществ в воздухе, а именно следующих химических элементов: Al, Ca, Cr, Cu, Fe, K, Mn, Ni, Pb, Sb, Sn, Ti, Zn. Также была получена информация о среднем возрасте умерших для каждого квадрата. Если предположить, что загрязнение воздуха вносит большой вклад в сокращение продолжительности жизни, а также, что другие негативные воздействия в среднем рассредоточены равномерно по городу, то, используя нейросеть, можно выделить влияние химического состава человека на исследуемую величину.

Нейронная сеть не накладывает ограничений на зависимость выходных данных от входных, таким образом, при выполнении упомянутых выше предположений вместо или наряду с информацией о воздухе можно исследовать влияние других факторов на продолжительность жизни, например, экономических и социальных. Однако большое количество входных параметров при небольшом объеме обучающего множества сделает обучение сети невозможным [2]. В документации к пакету Deductor приводится следующее правило для выбора числа нейронов в скрытом слое: «количество связей между нейронами должно быть примерно на порядок меньше количества примеров в обучающем множестве». Если рассмотреть сеть с одним скрытым слоем и одним выходом, то количество связей в ней ( $N_{link}$ ) задается выражением:

$$N_{link} = N_{input} * N_{hidden} + N_{hidden} ,$$

где  $N_{input}$  – количество входов,  $N_{hidden}$  – количество нейронов в скрытом слое. Таким образом:

$$N_{hidden} < 0,1N/(N_{input} + 1),$$

где  $N$  – размер обучающей выборки. Количество дискретов, которые используются для тренировки сети, в случае г. Тулы около 300. Если в качестве входных параметров использовать концентрации химических веществ, то мы должны использовать  $N_{hidden} < 2,2$ . Сеть из 2-х нейронов не способна смоделировать зависимость выхода от входов и процесс обучения не сможет сойтись [3].

Непосредственное использование первичных данных в процессе обучения в данном классе проблем обладает еще одним недостатком. Значения концентрации химических веществ в воздухе отдельного города сильно коррелированы между собой. Это связано, во-первых, с небольшим набором типов источников загрязнения (транспорт, котельные, химические и металлоперерабатывающие предприятия), во-вторых, с относительной схожестью ими выбрасываемых элементов. Подобные корреляции сильно затрудняют обучение сети и интерпретацию результатов. Предварительное обнаружение этих зависимостей позволит выявить ограничения для построенной модели, а именно наборы элементов, влияние которых на человека невозможно различить, используя имеющийся набор данных без привлечения физиологических моделей организма. Данное ограничение сильно уменьшает круг задач, которые можно решать таким подходом. Однако мы в первую очередь интересуемся проблемой улучшения КЖ населения в конкретном регионе, в котором источники загрязнения выбрасывают химические вещества в пропорциях близких к концентрациям присутствующим в воздухе данного региона. Таким образом, даже если нейросеть не будет способна различить влияние отдельных химических веществ, то она будет способна предсказать изменение продолжительности жизни при уменьшении выбросов на одном из источников в конкретном регионе. А именно такой вопрос и стоит при принятии решения об улучшении КЖ.

Таблица 1

Результат факторного анализа загрязнения атмосферы г. Тулы (для наглядности веса факторов < 0,2 заменены символом «-»)

№	Элемент	Фактор 1	Фактор 2	Фактор 3	Фактор 4	Фактор 5
1	Al	0,25	0,75	-	-	-0,29
2	Ca	0,7	-0,43	-0,32	-	-
3	Cr	0,29	0,58	-0,25	-0,33	0,55
4	Cu	0,42	0,34	0,35	0,41	-0,23
5	Fe	0,31	-0,34	0,64	-0,53	-
6	K	0,63	0,26	-0,58	-	-
7	Mn	0,35	-	0,65	-0,59	-
8	Ni	0,55	-	-0,51	-0,37	-
9	Pb	0,72	-	0,52	0,23	-
10	Sb	-0,42	0,71	-	-	-
11	Sn	0,68	-	-0,24	-	-0,49
12	Ti	-	0,8	-	-	-
13	Zn	0,58	-	0,35	0,55	0,28

Для построения модели, которая может предсказать изменение продолжительности жизни при варьировании определенного набора факторов, мы использовали пакет Deductor. В процессе работы были проделаны следующие шаги.

- Данные по смертности и химическому составу воздуха импортированы из текстовых файлов в таблицы. Площадь города была разбита на 750 дискретов размером 500×500 метров.
- Две таблицы объединены по уникальному индексу, присвоенному каждому квадрату на карте города.
- При помощи фильтра были отброшены дискреты, в которых количество смертей было меньше 10, так как они не несут статистически достоверной информации. В результате обработки осталось 238 записей пригодных для дальнейшего анализа.
- Проведен факторный анализ по 13 характеристикам химического состава воздуха: Al, Ca, Cr, Cu, Fe, K, Mn, Ni, Pb, Sb, Sn, Ti, Zn. Оставлены 5 факторов, которые объясняют 82% дисперсии исходных данных (таблица 1).
- Используя 5 факторов в качестве входов и средний возраст в момент смерти, обучена нейросеть с одним скрытым слоем и 5 нейронами.
- Полученная нейросеть может быть использована для предсказания изменения продолжительности жизни при изменении экологической обстановке в городе.

В общем случае интерпретация результатов факторного анализа затруднительна, поскольку факторы характеризуют дисперсию исходных случайных данных. Однако значения концентраций, рассматриваемые в данной работе, были измерены на регулярной сетке, покрывающей густонаселенные районы города. Таким образом, факторы характеризуют распределение вредных веществ в городе и зависят от расположения и типов источников загрязнения. Например, можно заметить, что веса фактора 1 (см. таблицу 1) для большинства элементов положительны и лежат в диапазоне 0,3-0,7, а его вклад составляет 25% от общей дисперсии. То есть можно считать, что фактор 1 характеризует некий средний уровень загрязнения воздуха в городе. Интерпретация остальных компонент не так очевидна и требует более сложного математического анализа.

Как уже было отмечено, использование факторов позволяет облегчить обучение нейронной сети, так как в процессе анализа исключаются все корреляции, присутствующие в исходных данных. Результаты тренировки сети с одним скрытым слоем и 5 нейронами приведены на рис. 1. Данные по 238 дискретам были разбиты на обучающее и тестовое множества в отношении 90% и 10% соответственно. Линейная независимость факторов позволила достичь распознавания 92% обучающих и 88% тестовых примеров. Визуально из рис. 1 можно заметить, что большинство результатов моделирования нейронной сетью лежат в диапазоне

3-4 года от исходных данных.

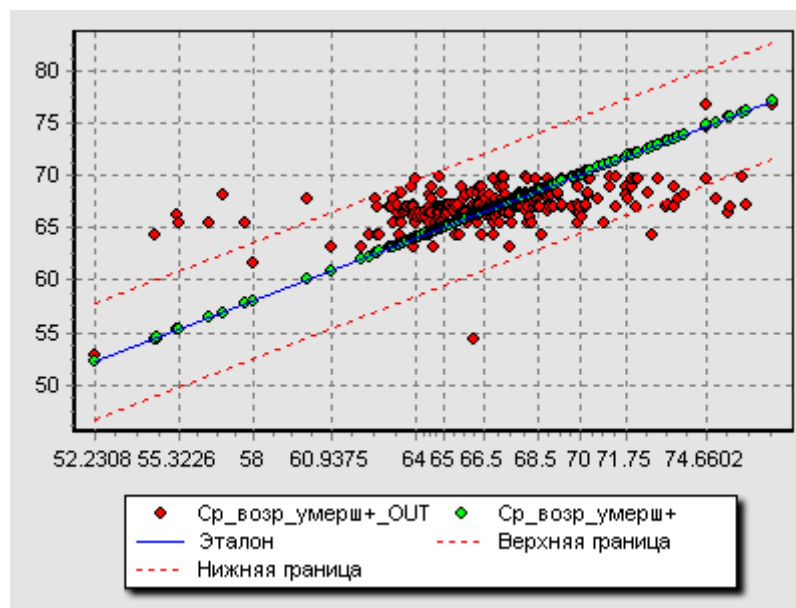


Рис. 1. Диаграмма рассеяния после обучения нейронной сети

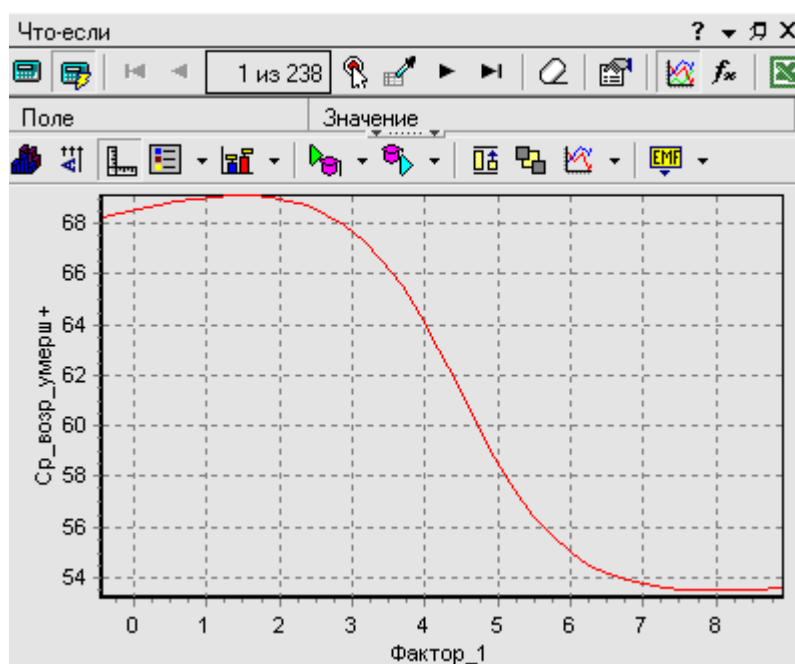


Рис. 2. График зависимости средней продолжительности жизни от фактора 1

Пакет Deductor позволяет построить зависимость выхода нейронной сети от одного из входов, результат для фактора 1, который соответствует общему уровню загрязнения в городе, приведен на рис. 2. Даже если учесть, что ошибка моделирования данной нейронной сетью составляет 3-4 года, средняя продолжительность жизни уменьшится с 68 до приблизительно  $54 \pm 3$  лет при увеличении концентрации. Значение фактора 1, равное 8, приблизительно соответствует 2-х кратному уровню загрязнения атмосферы от текущего среднего значения по городу. Влияние отдельных химических компонентов в рамках данного подхода на здоро-

вье человека требует дальнейшего более глубокого анализа и затруднительно средствами пакета Deductor.

Предложенный в данной работе подход к анализу влияния загрязнения атмосферы на среднюю продолжительность жизни может быть обобщен и на другие типы воздействий как экологические, так и социальные и экономические. Более того, используя в качестве выходного параметра нейросети частоту заболевания той или иной болезнью, можно построить модель для предсказания влияния различных факторов на здоровье населения.

### **Литература**

1. Машинцов Е.А., Кузнецов А.А., Лебедев А.М., Новосельцев В.Н., Математические модели и методы оценки экологического состояния территорий. – М.: Издательство физико-математической литературы, 2010.
2. James A. Freeman, David M. Skapura, Neural Networks Algorithms, Applications and Programming Techniques, Adison-Wesley, 1991.
3. Raul Rojas, Neural Networks A Systematic Introduction, Springer-Verlag, Berlin, 1996.



## **ПРИМЕНЕНИЕ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ В МОНИТОРИНГЕ И ОЦЕНКЕ ИННОВАЦИОННОЙ АКТИВНОСТИ ОРГАНИЗАЦИЙ ПРОМЫШЛЕННОСТИ**

*Мороз О.А., доцент, филиал «Взлет» федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Московский авиационный институт (национальный исследовательский университет)»*

В настоящее время мы являемся свидетелями перехода в постиндустриальную эпоху развития, которую также называют «информационным обществом», «обществом знаний» ввиду той роли, которую в ней играют знания и информация, а также «обществом услуг», так как в этом новом обществе все большая часть трудоспособного населения переходит на работу в сферу услуг [1].

В Концепции долгосрочного социально-экономического развития Российской Федерации на период до 2020 года подчеркивается, что с переходом в ближайшее десятилетие развитых стран к формированию новой технологической базы, основанной на использовании фундаментальных научных идей и достижений, прорывных технологий во всех системах жизнеобеспечения человека в целях устойчивого инновационного развития, целью является не стремление «догнать» какие-то страны, используя заимствованные у них методы и технологии, а достичь более высокого уровня развития, создать качественно новую социальную, экологическую, хозяйственную реальность. Такой представляется наиболее результативная линия [2].

Экономика знаний предъявляет высокие требования к профессиональной грамотности, знаниям умениям выпускников экономических специальностей.

В то же время, независимые опросы зарубежных исследователей показывают, что около половины (47%) компаний отметили, что не так легко найти и нанять квалифицированного инженера или техника. И для большинства компаний это вопрос наличия, а не вопрос стоимости.

Еще одной проблемой является качество образования. Из числа опрошенных руководителей, 35% оценивают качество университетского образования сегодняшних выпускников как низкое, по сравнению с 41% тех, кто склонен говорить, что этот уровень качества является высоким.

Из числа опрошенных, 51% оценили уровень выпускников профессионально-технических училищ и колледжей, как низкий и недостаточный для нужд своих компаний, в то время как положительные оценки дали только из 23% респондентов [3].

В сложившихся условиях, одной из стратегических составляющих устойчивого прогресса качества экономического образования становится освоение современной инструментальной базы, основанной, в том числе, на применении интеллектуальных информационных технологий.

Круг вопросов, решаемых в рамках применения современной технологии знаний, становится все более масштабным, расширяя не только горизонты восприятия традиционной аналитической информации для целей принятия решений, но и обогащая заинтересованных пользователей новыми формами ее получения.

В процессе обучения в рамках ряда дисциплин («Экономика отрасли», «Экономика предприятия», «Бизнес-планирование», «Прогнозирование», «Менеджмент исследований и разработок», «Инновационный менеджмент» и др.), учитывающих различные инновационные аспекты социально-экономического развития, возникает необходимость исследования различных аспектов инновационного развития отдельных субъектов, сегментов и рынков в целом, как одной из задач формирования целостного и системного восприятия происходящих социально-экономических процессов и развития навыков сбора, оценки и интерпретации экономической информации у обучаемых.

Расширение аналитического кругозора, практического и научно-теоретического базиса студентов также служит основой для развития навыков индивидуальной самостоятельной работы в процессе дипломного и курсового проектирования.

В связи с этим, определенный интерес представляет организация *мониторинга различных параметров инновационного развития организаций промышленности* в рамках активного освоения экономических образовательных дисциплин.

Росстат в установленном порядке осуществляет статистические наблюдения за инновационной деятельностью организаций промышленности и малых предприятий, используя следующую терминологию [4].

- *Организации, осуществлявшие технологические инновации* – организации, осуществлявшие разработку и внедрение новых или усовершенствованных товаров, работ, услуг, технологических процессов или способов производства (передачи) услуг и иные виды инновационной деятельности.
- *Инновационная активность* организации характеризует степень участия организации в осуществлении инновационной деятельности в целом или отдельных ее видов в течение определенного периода времени.
- *Уровень инновационной активности организаций* определяется как отношение числа организаций, осуществлявших технологические, организационные или маркетинговые инновации, к общему числу обследованных за определенный период времени организаций.
- *Технологические инновации* представляют собой конечный результат инновационной деятельности, получивший воплощение в виде нового либо усовершенствованного продукта или услуги, внедренных на рынке, нового либо усовершенствованного про-

цесса или способа производства (передачи) услуг, используемых в практической деятельности.

*Мониторинг* выступает в данном случае, как специально организованное наблюдение и отображение в определенных формах системной информации о различных параметрах инновационного развития организаций промышленности.

Современные системы мониторинга являются не только системами пассивной регистрации наблюдаемых явлений, но выступают как специальная контрольная среда, позволяя создать необходимую информационную систему поддержки управленческих решений, являясь также основой разработки перспективной политики в динамично изменяющихся условиях.

В качестве одного из направлений *мониторинга различных параметров инновационного развития организаций промышленности* выступает исследование воздействия результатов инновационной деятельности на развитие инновационно-активных организаций промышленности по направлениям инновационной деятельности.

**Цель исследования** – оперативное представление информационного образа воздействия результатов инновационной деятельности на развитие инновационно-активных организаций промышленности, характеризующегося определенным набором параметров в заданный момент времени.

**Задачи исследования** – сбор, аналитическая обработка, оценка и интерпретация инновационных процессов на промышленных предприятиях.

В качестве информационной базы использованы материалы государственных статистических наблюдений Росстата<sup>1</sup>.

Обобщение, оценка и интерпретация воздействия результатов инновационной деятельности на развитие организации в общем числе инновационно-активных организаций промышленности была реализована средствами технологической платформы для создания законченных аналитических решений *Deductor Academic*.

Аналитическая платформа *Deductor Academic* обладает широким арсеналом средств решения поставленных задач, которые характеризуются возможностями:

- импорта и хранения данных;
- обработки данных;
- визуализации информации.

В целях установления характеристик воздействия результатов инновационной деятельности на развитие инновационно-активных организаций промышленности исходные данные были подвергнуты интеллектуальному анализу с использованием самоорганизующихся *карт Конохена*.

---

<sup>1</sup> <http://www.gks.ru/>

Для этого последовательно осуществлялись шаги [6]:

- импорт данных в формате \*.txt;
- обработка данных с применением метода Data Mining в форме самоорганизующихся карт Кохонена;
- визуализация результатов.

Самоорганизующиеся карты могут использоваться для решения таких задач, как моделирование, прогнозирование, поиск закономерностей в больших массивах, выявление наборов независимых признаков и сжатие информации. Наиболее распространенное применение сетей Кохонена – решение задачи классификации без учителя, т.е. кластеризации [5].

Задача использования самоорганизующихся карт Кохонена в проведении мониторинге различных параметров инновационного развития сводится к использованию возможностей интеллектуального анализа для поиска закономерностей в массиве информации о воздействии результатов инновационной деятельности на развитие инновационно-активных организаций промышленности.

Импорт данных осуществляется последовательно в несколько этапов с использованием *Мастера импорта*. Процесс импортирования данных предполагает поэтапную настройку формата и других параметров, а также возможность способов отображения информации. Результат импорта данных представлен на рис. 1.

Результаты инновационной деятельности	высокая степень воздействия результатов инновационной деятельности на развитие производства	средняя степень воздействия результатов инновационной деятельности на развитие производства	низкая степень воздействия результатов инновационной деятельности на развитие производства	отсутствие воздействия результатов инновационной деятельности на развитие производства
Расширение ассортимента товаров, работ, услуг	43,2	38,1	12,1	29,8
Сохранение традиционных рынков сбыта	41,1	37	11,7	33,4
Расширение рынков сбыта	30,9	35,6	14,9	41,7
Улучшение качества товаров, работ, услуг	40,6	41,1	12	29,5
Замена снятой с производства устаревшей "продукции"	17,2	27,3	20,1	58,6
Увеличение занятости	13,8	30,4	28,4	50,6
Повышение гибкости производства	21,7	39,2	20,5	41,7
Рост производственных мощностей	27,6	32,7	19,9	42,8
Сокращение затрат на заработную плату	9,6	24	30,2	59,3
Сокращение материальных затрат	16	35	27,2	44,9
Повышение энергоэффективности производства (сокращение потребления энергии)	20,4	33,3	22,1	47,4
Улучшение условий и охраны труда	17,9	37,9	20,4	46,9
Сокращение времени на взаимодействие с клиентами или поставщиками	11	26,7	23,9	61,7
Повышение мотивации к осуществлению инновационной деятельности	12,2	31,8	26	53,2
Улучшение информационных связей внутри организаций или с другими организациями	16,9	32,6	21,1	52,6
Снижение загрязнения окружающей среды	13,7	31,9	23,6	54
Обеспечение соответствия современным техническим регламентам	35,3	36,4	15,7	35,9
Внедрение товаров, работ, услуг на новые рынки сбыта в новые группы стран	15,8	24,2	14,6	68,5
Внедрение товаров, работ, услуг на новые географические рынки	9,9	19,5	17,6	76,2

Рис. 1. Результат импорта данных

Исходные данные представлены в виде стандартной таблицы, и повторяют логику представления исходной информации.

С использованием *Мастера обработки*, производится соответствующая настройка параметров для обрабатываемой информации. Для

конкретного случая все поля, кроме поля заголовка, принимаются как входные, а в обучающее множество включаются все данные.

При выборе параметров кластеризации количество кластеров устанавливалось вручную.

На следующих этапах в автоматическом режиме определяются параметры карты Кохонена и параметры ее обучения.

После завершения процесса построения карты в настройках изображения карты Кохонена устанавливаются параметры ее отображения.

В результате кластеризации по степени оценки воздействия результатов инновационной деятельности на развитие организации в общем числе инновационно-активных организаций промышленности были выделены 5 кластеров (рис. 2).

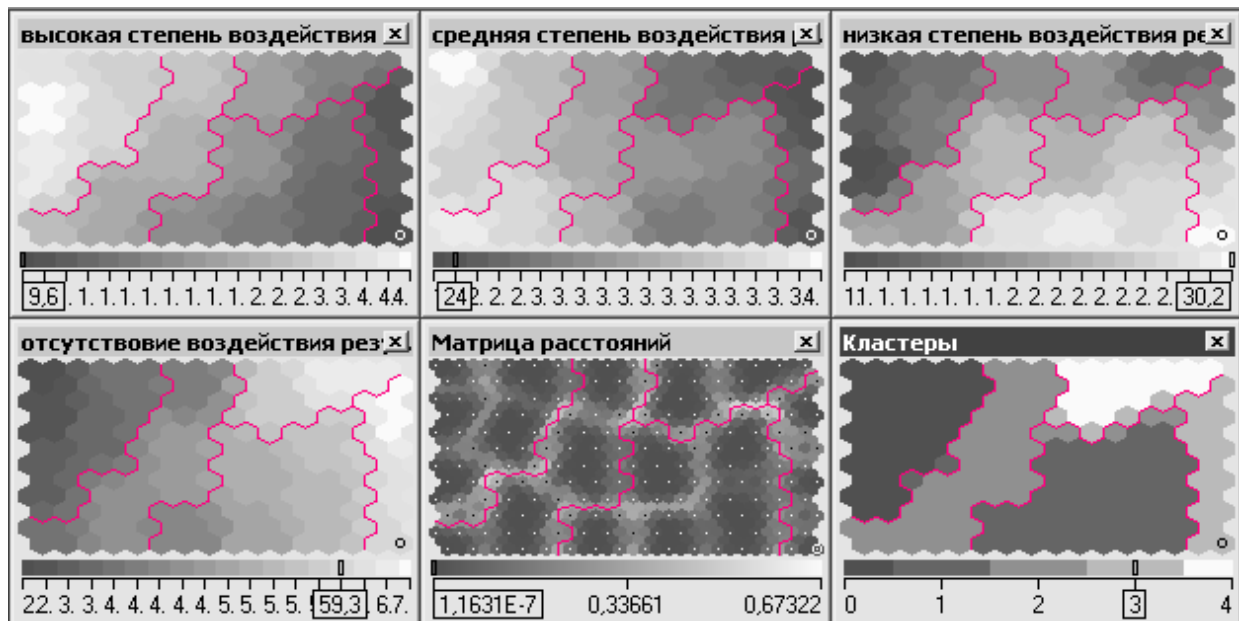


Рис. 2. Результаты кластерного анализа оценки воздействия результатов инновационной деятельности на развитие инновационно-активных организаций промышленности

Каждый из кластеров характеризуется определенным уровнем воздействия результатов инновационной деятельности на развитие промышленных организаций или их отсутствием в разрезе различных направлений и сфер этой деятельности.

Визуализированные результаты дают основания для многочисленных интерпретаций на предмет воздействия результатов инновационной деятельности на развитие инновационно-активных организаций промышленности в разрезе определенных направлений.

Дополнительные сведения о параметрах кластеров позволяют получить такие инструменты отображения, как матрица расстояний, матрица ошибок квантования, матрица плотности попадания и др.

Полученная в результате построения карт Кохонена группировка результатов представлена в таблице 1, которая составлена с использованием *Excel*.

Таблица 1

Результаты мониторинга и оценки воздействия результатов инновационной деятельности на развитие инновационно-активных организаций промышленности

Результаты инновационной деятельности	Степень воздействия результатов инновационной деятельности на развитие производства			
	высокая	средняя	низкая	отсутствие
<b>Кластер 1</b>				
Улучшение качества товаров, работ, услуг	40,6	41,1	12	29,5
Сохранение традиционных рынков сбыта	41,1	37	11,7	33,4
Расширение рынков сбыта	30,9	35,6	14,9	41,7
Расширение ассортимента товаров, работ, услуг	43,2	38,1	12,1	29,8
Обеспечение соответствия современным техническим регламентам, правилам и стандартам	35,3	36,4	15,7	35,9
<b>Среднее значение</b>	<b>38,22</b>	<b>37,64</b>	<b>13,28</b>	<b>34,06</b>
<b>Кластер 2</b>				
Улучшение информационных связей внутри организаций или с другими организациями	16,9	32,6	21,1	52,6
Увеличение занятости	13,8	30,4	28,4	50,6
Сокращение материальных затрат	16	35	27,2	44,9
Снижение загрязнения окружающей среды	13,7	31,9	23,6	54
Повышение мотивации к осуществлению инновационной деятельности	12,2	31,8	26	53,2
<b>Среднее значение</b>	<b>14,52</b>	<b>32,34</b>	<b>25,26</b>	<b>51,06</b>
<b>Кластер 3</b>				
Улучшение условий и охраны труда	17,9	37,9	20,4	46,3
Рост производственных мощностей	27,6	32,7	19,9	42,8
Повышение энергоэффективности производства	20,4	33,3	22,1	47,4
Повышение гибкости производства	21,7	39,2	20,5	41,7
<b>Среднее значение</b>	<b>21,9</b>	<b>35,76</b>	<b>20,73</b>	<b>44,7</b>
<b>Кластер 4</b>				
Сокращение затрат на заработную плату	9,6	24	30,2	59,3
Сокращение времени на взаимодействие с клиентами или поставщиками	11	26,7	23,9	61,7
Внедрение товаров, работ, услуг на новые географические рынки	9,9	19,5	17,6	76,2
<b>Среднее значение</b>	<b>10,17</b>	<b>23,4</b>	<b>23,9</b>	<b>65,73</b>
<b>Кластер 5</b>				
Замена снятой с производства устаревшей "продукции"	17,2	27,3	20,1	58,6
Внедрение товаров, работ, услуг на новые рынки сбыта в новые группы предприятий	15,8	24,2	14,6	68,5
<b>Среднее значение</b>	<b>16,5</b>	<b>25,75</b>	<b>17,35</b>	<b>63,55</b>

Использование статистики при выборе фильтра по кластеру дает возможность анализа каждого кластера. Визуализатор *Профили класте-*

ров обеспечивает представление статистических характеристик кластеров. Использование визуализаторов *Диаграмма* и *Гистограмма* также обеспечивается графическое представление результатов [6].

*Кластер 1* характеризуется результатами инновационной деятельности по тем направлениям, где достигнуты наиболее убедительные результаты, которые характеризуются убедительной степенью воздействия результатов инновационной деятельности на развитие производства по уровням «высокая» (более 38 %) и «средняя» (более 37 %). Степень «низкая» по этим направлениям фиксируется реже, чем положительный результат и его «отсутствие» (при рассмотрении средних значений).

Можно увидеть, что инновационная деятельность для этой совокупности была направлена на: «Улучшение качества товаров, работ, услуг», «Сохранение традиционных рынков сбыта», «Расширение рынков сбыта», «Расширение ассортимента товаров, работ, услуг» и «Обеспечение соответствия современным техническим регламентам, правилам и стандартам».

*Кластер 2* характеризуется менее удовлетворительными результатами по совокупности направлений инновационной деятельности. Достигнутые результаты встречаются более чем в 2 раза реже, чем в лучшей по этому критерию, выборке. В целом, степень воздействия характеризуется хорошими уровнем для «средней» степени воздействия инновационной деятельности на развитие производства, в значительной части, результаты характеризуют степень как «низкая» и даже «отсутствие» более, чем в 50 % случаев (также при рассмотрении средних значений).

Предметом приложения усилий в инновационной деятельности для этой совокупности являлись: «Улучшение информационных связей внутри организаций или с другими организациями», «Увеличение занятости», «Сокращение материальных затрат», «Снижение загрязнения окружающей среды», «Повышение мотивации к осуществлению инновационной деятельности».

*Кластер 3* характеризуется, преимущественно, «средними», «низкими» результатами и их «отсутствием». «Высокая» степень воздействия зафиксирована, в среднем, для 22% случаев.

В эту совокупность направлений инновационной деятельности попали: «Улучшение условий и охраны труда», «Рост производственных мощностей», «Повышение энергоэффективности производства», «Повышение гибкости производства».

*Кластер 4* обращает на себя внимание, в первую очередь, самым низким уровнем «высокой» степени воздействия результатов инновационной деятельности на развитие производства, в среднем, около 11% и самым высоким уровнем «низкой» степени воздействия результатов инновационной деятельности на развитие производства, в среднем, почти 66%. Около 24% внедрений дают «среднюю» или «низкую» степень воздействия.

Данный кластер представлен результатами в направлениях: «Сокращение затрат на заработную плату», «Сокращение времени на взаимодействие с клиентами или поставщиками», «Внедрение товаров, работ, услуг на новые географические рынки».

*Кластер 5* характеризуется самыми высокими уровнем «низкой» степени воздействия результатов инновационной деятельности на развитие производства, в среднем, порядка 64%. Только не более, чем в 16,5 % случаев фиксируется «высокая» степень воздействия, при этом «средняя» степень фиксируется чаще, чем «низкая» при ориентации на средние значения показателей.

Рассматриваемую совокупность составили такие направления, как «Замена снятой с производства устаревшей "продукции"» и «Внедрение товаров, работ, услуг на новые рынки сбыта в новые группы предприятий».

В целом, для инновационно-активных предприятий промышленности, можно отметить заметно недостаточную эффективность мероприятий, направленных на снижение себестоимости за счет более эффективного использования труда, проникновение на новые географические сегменты и сокращение цикла взаимодействия с потребителями и поставщиками, хотя частично усилия по сохранению и некоторому расширению рынков сбыта относительно результативны, равно, как и расширение ассортиментных линеек. Достаточно перспективными являются инновации, направленные на положительные качественные изменения результатов производственной деятельности, в том числе за счет обеспечения соответствия установленным уровням требований.

На основе полученных данных могут быть сделаны более детальные аналитические выводы для целей принятия решений по результатам мониторинга, оценки и интерпретации воздействия результатов инновационной деятельности на развитие инновационно-активных организаций промышленности, в том числе с использованием дополнительных информационные массивов.

Применение аналитической платформы *Deductor* позволяет проводить многоаспектные исследования с использованием различных инструментов обработки информации. В частности, аналитическая платформа позволяет решать задачи группировки и кластеризации с использованием самоорганизующихся карт Кохонена, обеспечивая выделение из совокупности объектов одинаковой природы некоторого количества относительно однородных групп с объединением их в кластеры.

При этом объединение объектов в группы (кластеры) происходит исходя из принципов соблюдения минимальных внутригрупповых и максимальных межгрупповых различий.

В результате достигается цель кластеризации (группировки) однородных объектов внутри их совокупности, что, в свою очередь, позволяет:

- выявить признаки, лежащие в основе группировки объектов;



- применять не только объектный, но и групповой подход к представлению информации, ее аналитическому использованию и интерпретации;
- агрегировать информацию без потери ее информационной ценности;
- использовать полученные результаты, как для принятия решений, так и для формирования новой информационной базы в практических и теоретических, научно-образовательных целях.

Приведенный пример демонстрирует значительный потенциал аналитической платформы *Deductor*, что определяет ее востребованность в решении широкого круга научно-образовательных задач обработки информации с использованием интеллектуальных методов анализа данных, в том числе при осуществлении мониторинга, оценки и интерпретации инновационной активности организаций промышленности. В результате внедрения и использования аналитических информационных систем в составе инструментальной образовательной базы достигается цель формирования у обучаемых системы теоретических знаний, практических умений и навыков для осуществления эффективной учебной и исследовательской деятельности.

## Литература

1. Тимофеева В.В. Методологические особенности организации инновационного образовательного комплекса [Электронный ресурс] / В.В. Тимофеева // Электронное научное издание «Устойчивое инновационное развитие: проектирование и управление». – 2010. – Том 6, №2. – Режим доступа: <http://www.gypravlenie.ru/?p=127>, свободный.
2. Концепция долгосрочного социально-экономического развития Российской Федерации на период до 2020 года, утв. Распоряжением Правительства Российской Федерации от 17 ноября 2008 г. № 1662-р // Собрание законодательства РФ. – 24.11.2008. – № 47, ст. 5489.
3. Prazdnichnykh A., Liuhto K. Can Russian companies innovate? - Views of some 250 Russian CEOs [Электронный ресурс] / Alexey Prazdnichnykh, Kari Liuhto // PEI Electronic Publications. – 21/2010. – Режим доступа: [http://www.tse.fi/FI/yksikot/erillislaitokset/pei/Documents/Julkaisut/Prazdnichnykh\\_Liuhto\\_final.pdf](http://www.tse.fi/FI/yksikot/erillislaitokset/pei/Documents/Julkaisut/Prazdnichnykh_Liuhto_final.pdf), свободный.
4. Указания по заполнению форм федерального статистического наблюдения № 2-наука «Сведения о выполнении научных исследований и разработок», № 3-информ «Сведения об использовании информационных и коммуникационных технологий и производстве связанных с ними товаров (работ, услуг)», № 4-инновация «Сведения об инновационной активности организации», утв. Приказом Росстата от 19.01.2009 № 4 // Вопросы статистики. – 2009. - № 6.

5. Data Mining: Учебное пособие / И. А. Чубукова. – М.: Интернет-Университет Информационных Технологий, 2008.
6. Паклин Н. Б. Орешков В.И. Бизнес-аналитика: от данных к знаниям (+CD): Учебное пособие. 2-е изд. – СПб.: Питер, 2010.

## **ИНФОРМАЦИОННАЯ МОДЕЛЬ ХРАНИЛИЩА ДАННЫХ КАФЕДРЫ ВУЗА**

*Бондарь И. А., доцент, Назарова С. А., преподаватель,  
Харьковский национальный экономический университет, г. Харьков*

В логической цепи, ключевыми звеньями которой выступают «уровень образования учащихся – уровень профессионализма работников – эффективность деятельности предприятия», одним из ведущих факторов выступает менеджмент такого высшего учебного заведения. В свою очередь, правильность управленческих решений на всех уровнях менеджмента вуза напрямую зависит от полноты учета, степени структуризации, достаточности предобработки и синхронизации данных, характеризующих оперативную внутреннюю и внешнюю деятельность такого рода организации, т.е. консолидации имеющейся информации, максимальная эффективность которой проявляется в составе информационно-аналитической системы вуза и его информационно-образовательной среды.

Нынешний этап развития программно-аппаратных средств автоматизации деятельности вуза (в составе которых обязательно имеют место инструменты хранения и консолидации информации) характеризуется следующим.

- Использование полифункциональных продуктов, таких как BPM (Business Performance Management, системы управления бизнес-процессами), ERP (Enterprise Resource Planning, системы управления ресурсами), которые охватывают практически все уровни управления вузом. Таким продуктом выступает система «Университет» компании «РЕДЛАБ» [1]. Основными особенностями инструментов хранения (в основном, это базы данных) данных в рамках подобных систем является выделенные для этого значительные мощности и универсальные для всех предприятий структуры метаданных, адаптация которых к полному учету оперативной информации о деятельности конкретного вуза требуют огромной методической работы, длительного и постоянного системного администрирования.
- Разработанность монофункционального инструментария для фиксации данных и консолидации информации по хозяйственным операциям, осуществляемым и фиксируемым такими подразделениями вуза, как касса, бухгалтерия, отдел материально-технического снабжения и обслуживания, редакционно-издательский отдел и другие обслуживающие основную деятельность вуза подразделения (например, редакционно-издательскими системы, системы бухгалтерского учета). Наиболее уязвимыми среди обслуживающих подразделений вуза вы-

ступают деканаты факультетов, обеспечивающие непосредственную организацию и контроль осуществления учебной деятельности. Для учета всех показателей оперативной учебной деятельности деканаты могут воспользоваться информационной системой баз данных «Электронный деканат», которая предполагает фиксацию, хранение и несложный анализ оперативных данных о ходе и результатах учебного процесса (студенческий контингент, успеваемость и т.д.) [2]. Данный вид инструментов хранения и консолидации данных предназначен только для освещения учебной деятельности с позиции конкретного структурного подразделения – деканата, т.е. имеет узкую предметную направленность.

Структурным подразделением, непосредственно осуществляющим основную (учебную и научную) деятельность вуза, является выпускающая кафедра. В силу большого разнообразия видов деятельности и современного состояния информатизации деятельности отечественных вузов нынешний уровень деятельности кафедры характеризуется:

- оперативные данные, обеспечивающие деятельность сотрудников кафедры (предоставляются другими подразделениями вуза) и описывающие ее результаты (внутренняя информация кафедры, которая предоставляется кафедрой для других подразделений) создается и хранится как в электронном, так и печатном виде;
- отсутствием полноценного и надежного инструментария учета и консолидации входной и исходящей информации кафедры.

Представленные обстоятельства обусловили цель данного исследования, которая состоит в разработке инструментария консолидации данных кафедры как основного элемента информационно-образовательной среды вуза, обеспечивающего комплексную автоматизацию всех видов деятельности кафедры и повышение эффективности использования учебно-методических информационных ресурсов на основе современных информационных и коммуникационных технологий.

Конкретной формой реализации указанного инструментария выступит хранилище данных кафедры, так как его основное преимущество по сравнению с аналогичными формами (базой данных) позволяет обеспечить не только целостность, но и непротиворечивость данных, высокую скорость и широкий диапазон методов анализа хранящихся данных, их достаточную для конкретных групп пользователей избирательность.

Основным функциональным назначением хранилища данных кафедры является своевременное предоставление необходимой для принятия эффективных управленческих решений консолидированной информации. Кроме того, в составе информационно-образовательного пространства вуза хранилища кафедр будут интегрированы в виртуальные кафедры, которые также предназначены для повышения качества и открытости образования, перехода к открытому непрерывному образованию, обеспечение конкурентоспособности высших учебных заведений на

рынке образовательных услуг.

Для обеспечения полноты и удовлетворительного (для целей дальнейшего управления) требуемого уровня консолидации данных необходимо разработать информационную модель, которая бы описывала поведение объекта исследования – кафедры и позволяла хранить имеющуюся и накапливать актуальную информацию о процессах кафедры. Решению данной научной задачи посвящена эта статья.

Для построения модели деятельности кафедры целесообразно использовать деятельностный и процессный подходы. Применение деятельностного подхода позволит выделить укрупненные направления деятельности кафедры.

При использовании процессного подхода в рамках каждого из укрупненных направлений деятельности кафедры возможно выделение бизнес-процессов из которых оно состоит.

В свою очередь, бизнес-процессы рассматриваются как информационные потоки, которые принимают информацию на входе и генерируют новую информацию на выходе. В ходе выполнения бизнес-процесса должна аккумулироваться конкретная информация, описывающая поведение процесса в рамках каждого из видов деятельности. Различные бизнес-процессы находятся в непрерывном взаимодействии, постоянно обмениваются информацией (между ними циркулируют информационные потоки) [3].

Для создания (описания) полной информационной модели деятельности кафедры (обобщенное форматизированное описание приведено в [4]) проведен анализ осуществляемых ее сотрудниками бизнес-процессов, а также бизнес-процессов, которые они обеспечивают. Данный анализ имел теоретический и эмпирический характер.

Вначале на основании анализа имеющихся в литературе данных выделен ряд направлений деятельности кафедры как основного подразделения вуза, задействованного во всех его бизнес-процессах.

Деятельность кафедры – совокупность всех видов работ по реализации образовательного процесса. Условно деятельность кафедры можно разделить на учебную и внеучебную профессиональную. Учебная деятельность кафедры заключается в осуществлении преподавания общих и специальных курсов, организации самостоятельной работы студентов, руководстве курсовыми и выпускными квалификационными работами и практиками. Под внеучебной профессиональной деятельностью кафедры вуза понимается совокупность всех видов деятельности кафедры, обеспечивающих образовательный процесс [4].

Таким образом, определены (согласно [5]) четыре основных вида деятельности кафедры: учебная (основная, педагогическая), методическая, научно-исследовательская и организационная.

Причем, уже теоретический анализ показал, что указанные виды деятельности целесообразно рассматривать в двух аспектах – стратегическом и оперативном, так как информация, характеризующая первый

аспект какого-либо направления деятельности, существенно отличается (хотя и тесно взаимосвязана) от той, которая характеризует оперативный аспект этого же вида деятельности.

Кроме того в результате теоретического анализа нормативных документов, регламентирующих деятельность структурных подразделений и вуза в целом установлено место кафедры в организационной структуре вуза.

Синтез результатов анализа места кафедры в организационной структуре вуза с итогами определения функциональных прав и обязанности кафедры позволил выявить направления и содержание информационных потоков выпускающей кафедры, условно представленных на рис. 1.



Условные обозначения связей (виды входящей и исходящей информации):

- 1 – учебные планы, разрабатываемые на основе утвержденных профильным министерством норм, учитывающие государственные, региональный и вузовские компоненты, условия и особенности обучения различных групп учащихся;
- 2 – рабочие планы учебных дисциплин;
- 3 – результаты текущего модульного и итогового контроля знаний студентов по дисциплине;
- 4 – списки профессорско-преподавательского состава кафедры;
- 5 – результаты работы государственных экзаменационных комиссий;
- 6 – численность и форма обучения зачисленных студентов на обучение на выпускаемые кафедрой специальности;
- 7 – расписание экзаменов;
- 8 – план мероприятий профессиональной ориентации в других учебных заведениях;
- 9 – результаты научно-исследовательской работы студентов;
- 10 – результаты выполнения индивидуальных заданий по учебной дисциплине;
- 11 – результаты комплексного курсового проектирования;
- 12 – результаты дипломного проектирования;
- 13 – заполненные по окончанию бакалавриата и магистратуры зачетные книжки;
- 14 – график консультаций профессорско-преподавательского состава кафедры;
- 15 – перечень конференций (условия их проведения), в которых могут принять участие студенты и сотрудники кафедры;
- 16 – план изданий кафедрой методической литературы;
- 17 – тема, план и результаты выполнения научно-исследовательской работы про-

фессорско-преподавательским составом кафедры;

- 18 – объемы и структура запланированной учебной нагрузки по дисциплинам, которые преподают сотрудники кафедры
- 19 – командировочные листы (для переноса занятий командированного в расписании) сотрудников кафедры;
- 20 – график учебного процесса;
- 21 – расписание аудиторных занятий;
- 22 – личные данные студентов (для кураторов);
- 23 – сводные данные текущего модульного контроля по всем дисциплинам (для кураторов);
- 24 – методическое обеспечение учебных дисциплин кафедры, учебных пособий, учебно-методических материалов по проведению всех видов занятий, а также других пособий, предусматривающих использование новых форм и методов преподавания;
- 25 – данные об уровне подготовки и участии в мероприятиях повышения квалификации педагогических кадров;
- 26 – заявки на проведение кафедрой научно-практических конференций;
- 27 – отчеты аспирантов и стажеров;
- 28 – рабочие планы профессорско-преподавательского состава кафедры;
- 29 – рабочие журналы профессорско-преподавательского состава кафедры;
- 30 – заявки на необходимые для обеспечения учебного процесса программно-аппаратные средства;
- 31 – контракты и заявления сотрудников кафедры о трудоустройстве в вузе;
- 32 – заявки на приобретение программного обеспечения, техники и других материальных средств, обеспечивающих организацию и проведение аудиторных занятий;
- 33 – рукописи методических материалов профессорско-преподавательского состава кафедры после правки редактора и корректора;
- 34 – расчетные листы по оплате труда;
- 35 – списки студентов академических групп.

Рис. 1. Прямые информационные потоки кафедры вуза

При этом детально изучено прямое информационное взаимодействие кафедры, т. е. движение информации между персоналом кафедры и ее непосредственным, самым близким окружением. Такая позиция авторов при анализе теоретического материала обоснована тем, что основные параметры информационного взаимодействия прямых партнеров кафедры с другими (не с кафедрой) подразделениями и отдельными должностными лицами вуза, которые, однако, касаются деятельности кафедры, напрямую обязательно учитываются при прямом взаимодействии с кафедрой.

Интенсивность указанных на рис. 1 информационных потоков изучена в ходе длительного эмпирического анализа – наблюдения за деятельностью персонала кафедры компьютерных систем и технологий (выпускающей для таких специальностей, как «Технологии мультимедийных издательств» и «Компьютерные технологии и системы издательско-полиграфических производств») Харьковского национального экономического университета, анализа документации этой кафедры, ее электронной информационной базы и движения информации.

В ходе эмпирического анализа деятельности конкретной кафедры установлено, что кроме бизнес-процессов, в которых используется обозначенная на рис. 1 исходящая и входящая информация, на кафедре реализуются внутренние процессы, обеспечиваемые такими внутренними данными как концепция развития кафедры (36), распределение организационных полномочий (37), графики и вопросы заседаний кафедры (38),

графики взаимных посещений (39), распределение студентов между руководителями – преподавателями кафедры по диплому магистра (40), специалиста (41), бакалавра (46) по научно-исследовательской работе (42), графики, регламентирующие дипломное проектирование магистров (43), специалистов (44), бакалавров (45) и курсовое проектирование (47), данные о базах производственной практики студентов (48).

Данные, представленные на рис. 1, показывают все многообразие входной и исходящей информации кафедры, однако не указывают на то, каким образом связаны данные, движущиеся в различных информационных потоках, что позволило бы вскрыть имеющуюся организацию информации на кафедре (построить информационную модель данных кафедры), при необходимости усовершенствовать ее и начать проектирование хранилища данных кафедры.

На основании выделения семантических связей между указанными на рис. 1 видами входящей и исходящей информации авторами статьи построена информационная модель организации данных кафедры в виде ориентированного сетевого графа (рис. 2).

Представленная на рис. 2 информационная модели организации данных кафедры позволяет:

- определить существующую семантическую взаимосвязь (еще не обобщенную и не консолидированную в электронном виде на кафедре) между выделенными на рис. 1 информационными потоками;
- определить ключевые узлы информации, каждый из которых будет представлен в виде таблиц измерений в рамках хранилища данных кафедры;
- установить обособленные данные, которые не взаимосвязаны с ключевыми узлами имеющейся на кафедре информации (например, при составлении плана изданий кафедрой методической литературы напрямую не учитываются объемы и структура запланированной учебной нагрузки по дисциплинам и объемы запланированной научно-исследовательской работы каждого из представителей профессорско-преподавательского состава), соответствующим образом скорректировать взаимосвязи таких данных;
- при необходимости (после изменения взаимосвязей между видами входящей и исходящей информации) переопределить ключевые узлы информации кафедры.



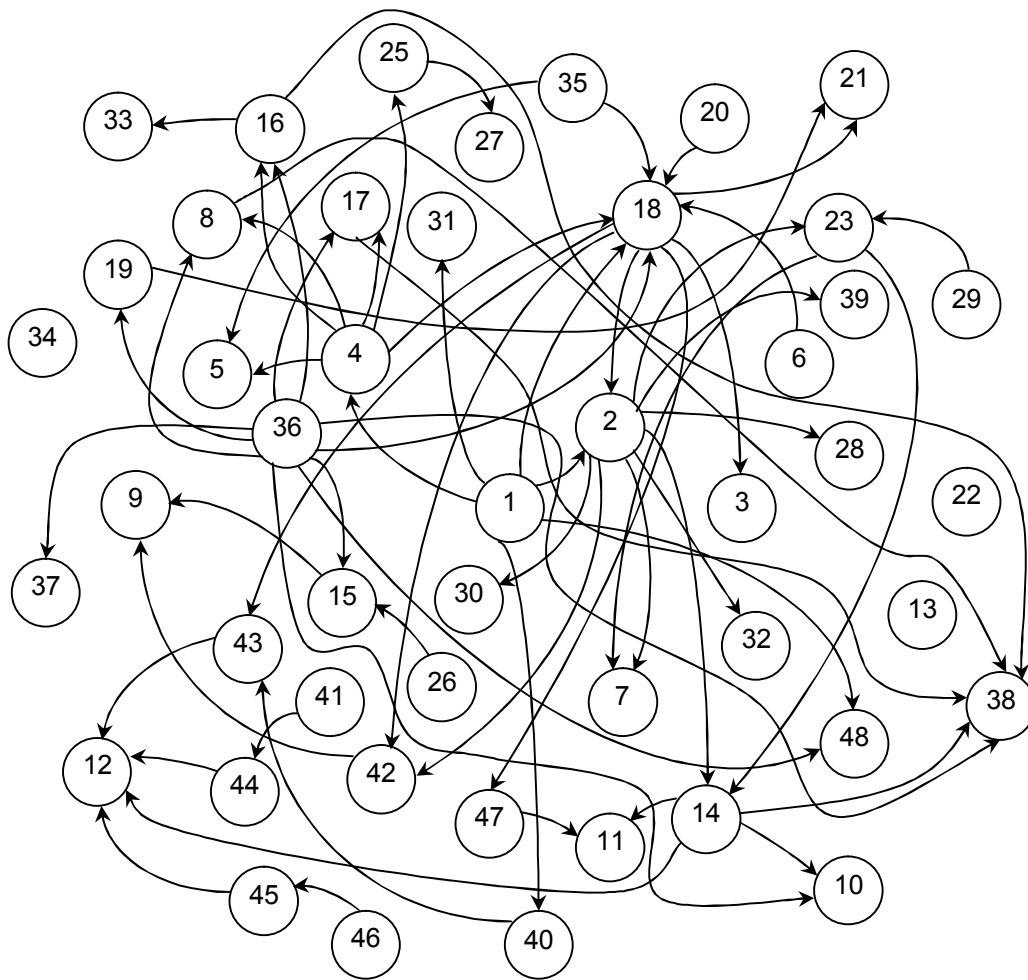


Рис. 2. Информационная модель организации данных кафедры

Указанные изменения преобразуют информационную модель организации данных в информационную модель хранилища данных кафедры.

С точки зрения авторов статьи, рациональная практическая реализация разработанной информационной модели хранилища данных кафедры предполагает использование:

- аналитической платформы Deductor (как программного инструментария, дающего полную и удобную возможность проектировать хранилища данных и выполнять все необходимые операции с размещенными в нем данными) для проектирования и наполнения хранилища данными;
- СУБД «FireBird» (как свободно распространяемой СУБД, системные возможности которой соответствуют сложности проектируемого хранилища данных) в целях хранения данных.

Кроме формирования в структуре метаданных основных таблиц (для измерений и процессов) в процессе создания хранилища данных планируется создание кубов данных, срезы по которым будут содержать наиболее часто востребованные наборы данных.

На завершающем этапе проектирования данных (при обеспечении локального и глобального доступа к нему в составе информационно-

образовательной среды вуза) планируется разработка витрин данных для каждой из групп (и подгрупп), заинтересованных в хранящейся и обработанной информации, которыми являются:

- сотрудники кафедры (функциональные подгруппы: заведующий кафедрой, кураторы специальностей, кураторы групп, консультанты по курсовому и дипломному проектированию и т.д.);
- студенты (функциональные подгруппы организованы по курсам и факультетам);
- сотрудники деканатов (функциональные подгруппы организованы по факультетам);
- сотрудники других обеспечивающих подразделений вуза (функциональные подгруппы: научно-исследовательский отдел, методический отдел, учебный отдел, бухгалтерия, издательство, отдел кадров и т. д.)
- внешние (по отношению к вузу) партнеры кафедры (функциональные подгруппы: школы, техникумы и лицеи, кафедры, выпускающие аналогичные специальности в других вузах и т. д.).

Представленная в данной статье логика всестороннего и детального анализа деятельности кафедры вуза позволила построить информационную модель хранилища данных как инструмента консолидации и обработки имеющейся информации с целью обеспечения правильности принимаемых управленческих решений и повышения открытости информационно-образовательной среды вуза.

Следующий этап исследований авторов данной статьи состоит в проектировании хранилища данных кафедры компьютерных систем и технологий Харьковского национального экономического университета средствами аналитической платформы Deductor Studio и СУБД «FireBird», результаты чего будут представлены на будущих конференциях «Бизнес-аналитика. Использование аналитической платформы Deductor в деятельности учебных заведений».

### **Литература**

1. Иевенко М.В., Смелянский Р.Л. Возможности системы «Университет» компании «РЕДЛАБ» для формирования и реализации стратегии развития вуза [Электронный ресурс]. Режим доступа: <http://www.umj.ru/index.php/pub/inside/436/>.
2. Аналитические возможности информационной системы баз данных «Электронный деканат». [Электронный ресурс]. Режим доступа: <http://www.mgimo.ru/about/structure/3495/3699/docs/63506/>
3. Кузнецов А. И. Методика проведения обследования бизнес-процессов компании // Технологии корпоративного управления. Информационные технологии [Электрон. ресурс]. Режим доступа:

[http://www.iteam.ru/publications/it/section\\_51/article\\_1469](http://www.iteam.ru/publications/it/section_51/article_1469).

4. Климович. Л. А. Показатели деятельности кафедры на основе векторных оценок // Университетское управление: практика и анализ. – № 3. – 2008. – С. 73-77.
5. Лемешева Т.Л. Виртуальные кафедры и инструментальные средства их проектирования // Научная сессия МИФИ-2003 г. Сборник научных трудов. Интеллектуальные системы и технологии. – М.: МИФИ, 2003. – С. 189-190.

## **ИНФОРМАЦИОННО-АНАЛИТИЧЕСКАЯ СИСТЕМА МОНИТОРИНГА СОСТОЯНИЯ ВОДНЫХ РЕСУРСОВ НИЖЕГОРОДСКОГО РЕГИОНА**

*Прокопенко Н.Ю., доцент, Родионова С.В., старший преподаватель,  
Чубченко М.Н., студент, Нижегородский государственный архитек-  
турно-строительный университет, г. Нижний Новгород*

В настоящее время проблема загрязнения рек, озер, водохранилищ, подземных вод, тесно связана с проблемой обеспеченности пресной водой, поэтому наблюдениям и контролю за уровнем загрязнения водных объектов уделяется особое внимание. Для осуществления этих функций проводится экологический мониторинг воды, который представляет собой систему наблюдений, оценки и прогноза изменений состояния водных объектов и окружающей среды в целом. Он включает:

1. регулярные наблюдения за состоянием водных объектов, количественными и качественными показателями состояния водных ресурсов, а также за режимом использования водоохранных зон;
2. сбор, обработку и хранение сведений, полученных в результате наблюдений;
3. внесение сведений, полученных в результате наблюдений, в государственный водный реестр;
4. оценку и прогнозирование изменений состояния водных объектов, количественных и качественных показателей состояния водных ресурсов.

Информация о состоянии водных ресурсов широко используется в различных сферах деятельности. Она необходима для:

1. своевременного выявления и прогнозирования развития негативных процессов, влияющих на качество воды в водных объектах и их состояние, разработки и реализации мер по предотвращению негативных последствий этих процессов;
2. оценки эффективности осуществляемых мероприятий по охране водных объектов;
3. информационного обеспечения управления в области использования и охраны водных объектов, в том числе в целях государственного контроля и надзора за использованием и охраной водных объектов.

Информационно-аналитическая система мониторинга состояния водных ресурсов предназначена для автоматизации функций отдела

учета объектов негативного воздействия и экологического мониторинга Министерства экологии и природных ресурсов Нижегородской области.

Анализ деятельности этого отдела позволил определить основные проблемы, требующие оптимизации процесса проведения мониторинга.

1. Данный процесс содержит огромный объем сопровождающей документации, что порождает проблемы:
  - поиска необходимой документации;
  - существует вероятность потери документов;
  - большие стоимостные затраты на покупку бумаги;
  - архивы документов сокращают дефицитную офисную площадь.
2. Наличие очень сложных процессов обработки и анализа данных, обусловленных необходимостью:
  - постоянно совершать анализ данных;
  - постоянной сверки данных с уровнем предельно допустимой концентрации и с показателями за предшествующий период.
3. Необходимость предоставления администрации региона точной, своевременной и актуальной информации в удобном для восприятия виде, т.е. в формате графиков, диаграмм, сводных таблиц и отчетов.

Для моделирования основных бизнес-процессов отдела учета объектов негативного воздействия и экологического мониторинга Министерства экологии и природных ресурсов Нижегородской области с целью оптимизации деятельности этого отдела были решены следующие задачи:

- смоделирован процесс проведения мониторинга водных ресурсов в модели IDEF0;
- смоделирован процесс «Считывание и обработка данных» «As is», а также процесс «Анализ данных» «As to be» в модели IDEF3;
- смоделированы основные потоки информации организации при проведении мониторинга водных ресурсов в модели DFD.

В ходе моделирования бизнес-процессов организации были выявлены так называемые «узкие места» функций процесса. В данном случае к «узким местам» были отнесены функции обработки, анализа и хранения данных (эти функции в настоящее время реализуются с помощью персонального компьютера, программы MS Excel), а также составление отчетности на протяжении проведения мониторинга, поскольку реализация данных функций требует значительных материальных и временных затрат. Таким образом, был сделан вывод,

что автоматизация этих функций позволит ускорить процесс проведения мониторинга, а также обеспечить высокое качество его проведения.

Оптимизация процесса проведения мониторинга наиболее эффективна с помощью создания информационной системы средствами СУБД MS Access и АП Deductor Studio. Выбор MS Access 2007 обоснован требованиями заказчика о замене существующего способа хранения и обработки данных установленным ПО (MS Office 2007) на пользовательских компьютерах в отделе. Главным критерием выбора аналитической платформы Deductor для разработки автоматизированной системы мониторинга среди других аналитических систем является наличие в АП Deductor самых современных методов извлечения, манипулирования, визуализации данных, технологий интеллектуального анализа, таких, как кластеризация, дерево решений, нейронные сети.

На основе выполненного проектирования была реализована в MS Access база данных, а именно были созданы необходимые таблицы, запросы, формы и отчеты, был разработан интерфейс автоматизированной системы. Структура базы данных «Мониторинг» для показателей загрязнения водных ресурсов разработана в соответствии со стандартами проектирования. Были выделены основные сущности и атрибуты, определены первичные ключи, была создана инфологическая модель, даталогическая модель, представлена схема данных. Основное назначение БД – сбор данных. На этапе программной реализации базы данных были созданы многочисленные запросы и на их основе отчеты по требованию заказчика.

Создание интегрированного хранилища данных «Мониторинг» было реализовано на базе Deductor Warehouse. Структура ХД представлена на рис. 1. При данной структуре ХД предполагается, что уникальность точки в пространстве определяется совокупностью измерений *Номер записи + Дата + Код объекта + Код загрязнителя + Створ*. Исходными данными для наполнения ХД служили 3 файла: **Измерения.txt**, **Водные объекты.txt** и **Загрязнители.txt**.

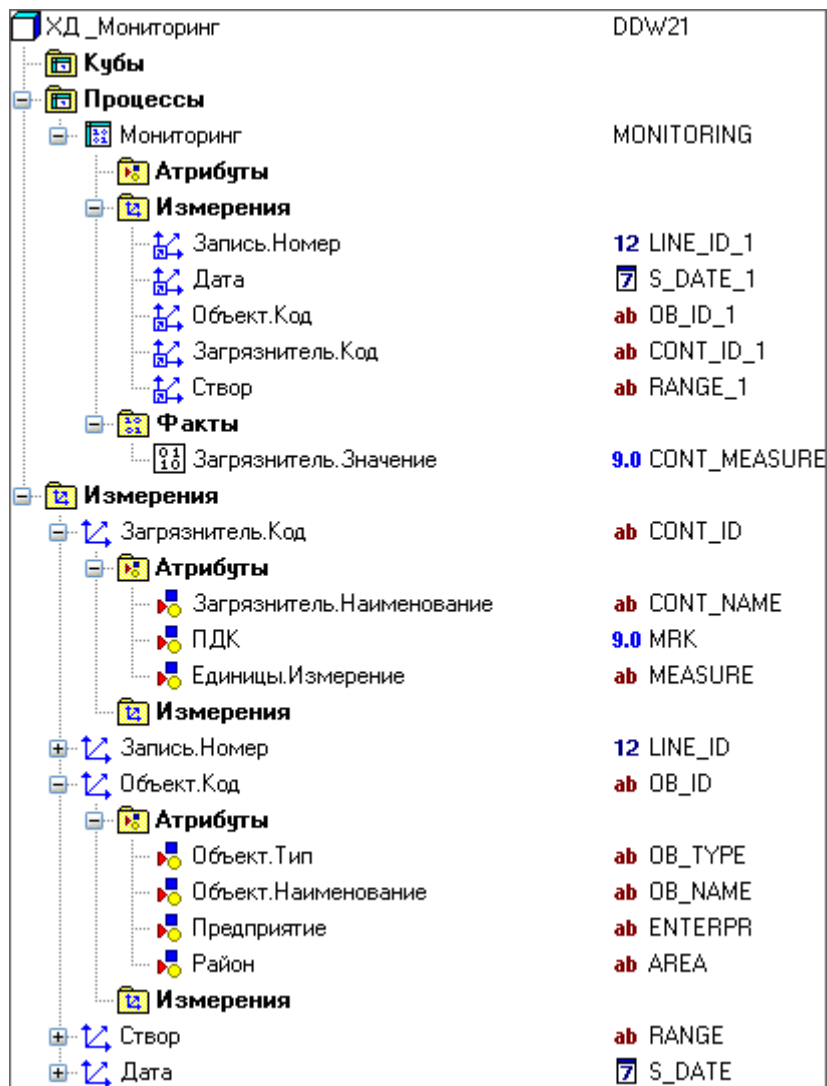


Рис. 1. Структура ХД «Мониторинг»

Целью обработки данных в Deductor является предоставление информации в удобном для восприятия и анализа виде. Используя возможности OLAP-анализа, были получены многочисленные отчеты в виде OLAP-кубов и кросс-диаграмм, которые отражают динамику загрязнения водных объектов на протяжении всего периода загрязнения, отношение того или иного загрязнителя к уровню его предельно допустимой концентрации (ПДК), процентное соотношение качества вод региона и другие показатели.

Построение некоторых отчетов требовало написания сценария в Deductor. Так для составления отчета «Индекс загрязнения воды (ИЗВ)» для всех рассматриваемых водных объектов была произведена фильтрация по каждому водному объекту и шести загрязнителям, относящимся к конкретному водному объекту. С помощью внешнего левого соединения для каждого водного объекта была составлена таблица, содержащая информацию о загрязнителях, их значениях и ПДК. Далее с помощью калькулятора в таблицу было добавлено поле «Отношение», которое вычис-

лялось путем деления столбца «Значение загрязнителя» на столбец «ПДК». После вычисления данного столбца была произведена группировка по измерениям «Наименование объекта» и «Дата анализа», а в качестве агрегации факта «Отношение» была выбрана сумма. Затем с помощью калькулятора в полученную таблицу было добавлено поле «ИЗВ», которое было вычислено по заданной формуле. Анализ полученной на основе этого отчета диаграммы позволил определить самый загрязненный объект из рассматриваемых – озеро Спасское.

Отчет «Отношение к ПДК», в котором все имеющиеся величины представлены в относительном виде был использован для проведения корреляционного анализа показателей. В ходе корреляционного анализа необходимо было выявить связи между загрязнителями или убедиться в их отсутствии на примере водного объекта «Озеро Нижневыксунское». С помощью обработчика «Корреляционный анализ» были получены матрицы корреляции, в которых выходными параметрами поочередно были каждый из загрязнителей исследуемого водного объекта. Наибольшая корреляционная зависимость была выявлена между показателями «нефтепродукты» и «ХПК», что объясняется повышенным химическим потреблением кислорода при наличии в воде повышенного содержания нефтепродуктов, в несколько раз превышающих ПДК.

Для отнесения водных объектов на основе рассчитанного выше показателя ИЗВ к определенному классу вод была построена модель Дерево решений, с помощью которой были получены правила отнесения того или иного водного объекта к конкретному классу вод.

Используя имеющуюся в хранилище данных информацию о мониторинге водных объектов Нижегородской области за 3 года: с 2008 по 2010 год, была построена модель прогноза динамики загрязнения выбранного водного объекта тем или иным загрязнителем на основе метода выделения компонент временного ряда. С помощью инструментов Deductor Studio «Настройка набора данных» и «Фильтр» были выбраны необходимые данные для построения прогноза, а именно был выбран водный объект «озеро Нижневыксунское» и загрязнитель «ХПК». Для проверки результатов прогноза в модель не было включено последнее из имеющихся значений загрязнителя. Точность модели оценивалась по ошибке МАРЕ, которая составила 1,6 %, что говорит о хорошем качестве построенной модели. После построения прогноза значение загрязнителя за 01.12.2010 год было сравнено с имеющимся реальным значением на ту же дату. Расхождение прогнозируемого значения с фактическим составило 0,9, что составляет 3,9 % и говорит о довольно точном результате построенной модели прогноза.



С помощью АП Deductor также были построены несколько нейросетевых моделей прогноза.

Таблица 1

Конфигурации нейронной сети

Параметры	Значение
Количество нейронов во входном слое	1; 2; 3
Количество скрытых слоев	2
Количество нейронов в скрытых слоях	2
Выходной слой	1
Тип активационной функции	Сигмоида

После построения прогноза значение загрязнителя за 01.12.2010 год было сравнено с имеющимся реальным значением на ту же дату. Расхождение прогнозируемого значения с фактическим составило 0,234 (для первого случая), что составляет 1,02 %; 0,702 (для второго случая), что составляет 3,05 % и 0,28 (для третьего случая), что составляет 1,21 %.

Также в качестве критерия оценки точности построенных нейронных сетей были использованы диаграмма рассеяния и среднеквадратическая ошибка прогноза. Для нейросетевой модели с тремя нейронами во входном слое получились наиболее близкая к эталонному значению диаграмма рассеяния и наименьшая среднеквадратическая ошибка (12 %).

Для обеспечения автоматического доступа к инструментарию аналитической платформы Deductor Studio из СУБД MS Access 2007 был разработан специальный механизм интеграции, позволяющий аналитику непрерывно работать с информационной системой. Интеграция приложений выполнена в данной работе средствами MS Access (рис. 2). Главная форма организует запуск приложений путем использования встроенных макросов. Аналитическая платформа Deductor, в свою очередь, имеет возможность выполнения автоматизированных сценариев и готовых отчетов путем применения пакетной обработки. Данный механизм позволяет обеспечить автоматический запуск платформы с открытием панели отчетов необходимого сценария.

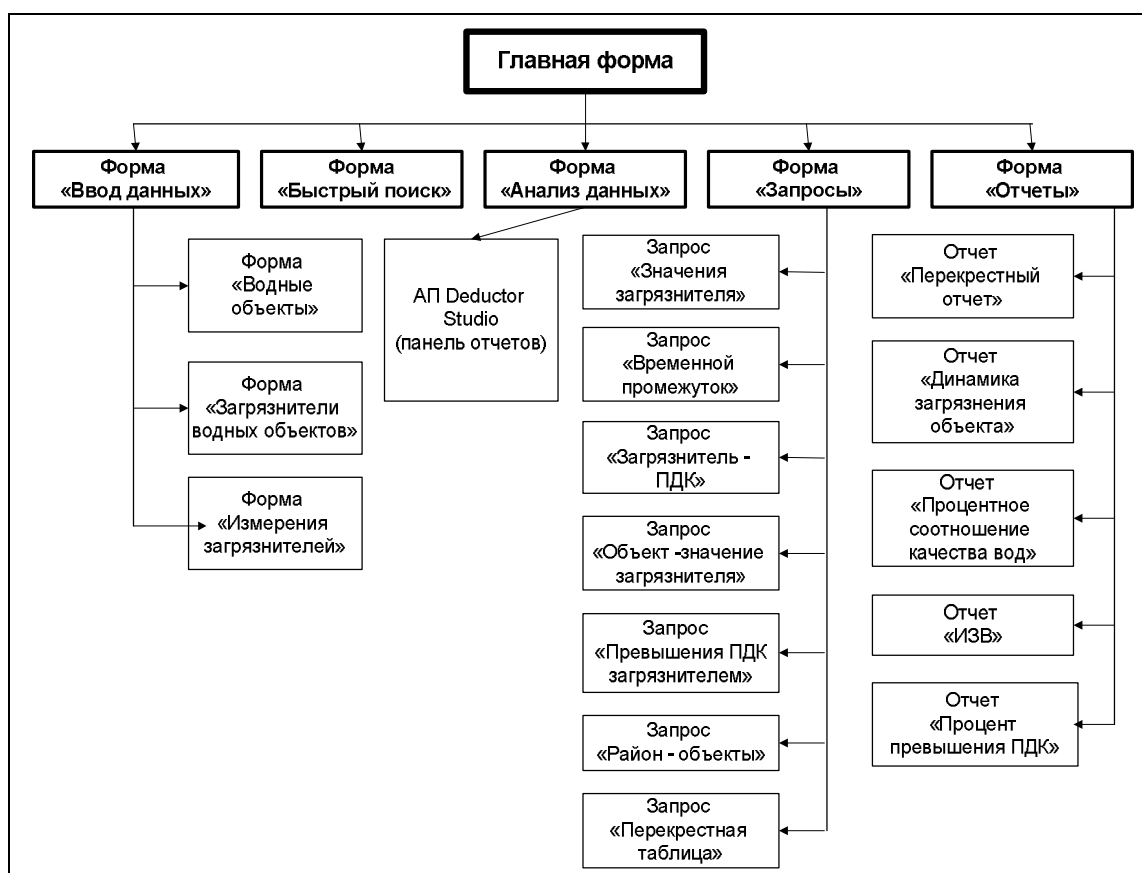


Рис. 2. Интеграция приложений в MS Access

После обеспечения интеграции был проведен расчет возможной экономической эффективности внедрения АИС в работу Министерства экологии и природных ресурсов. Для оценки эффективности разработанной автоматизированной информационной системы были составлены функциональные модели процесса проведения мониторинга водных ресурсов в системе ARIS. Были составлены диаграммы процессов проведения мониторинга водных ресурсов до и после оптимизации и проведена симуляция работы процессов. В результате сравнения полученной статистики было установлено, что использование разработанной АИС позволяет снизить временные затраты на проведение данного процесса более, чем в 5 раз за счет уменьшения времени, затрачиваемого на обработку, анализ данных и составление отчетности. Это позволяет освободить время для решения сотрудниками отдела экологического мониторинга других задач.

Таким образом, на основании проделанной работы, можно сделать вывод о том, что разработанная АИС для проведения мониторинга водных ресурсов региона позволит эффективнее организовать работу сотрудников Министерства экологии и природных ресурсов Нижегородской области, усилит взаимодействие между участниками проекта и впоследствии повысит эффективность проведения мониторинга водных ресурсов в Нижегородской области.

## **ИСПОЛЬЗОВАНИЕ АП DEDUCTOR ДЛЯ АНАЛИЗА РЕЗУЛЬТАТОВ И ПЛАНИРОВАНИЯ ПРОФИОРИЕНТАЦИОННОЙ РАБОТЫ ВУЗА**

*Рыбанов А.А., доцент, Зайчук О.А., студент,  
Волжский политехнический институт (филиал) ВолгГТУ, г. Волжский*

Конкурентоспособность и эффективное развитие института высшего образования обеспечивается путем непрерывного постоянного инновационного процесса, ключевые направления которого:

- внедрение новых информационных технологий;
- использование современных методов управления.

Повысить эффективность такой сложной системы, как ВУЗ, можно лишь путем адекватного регулирования многообразных связей между различными видами деятельности. Обеспечение руководства аналитической информацией позволит управлять вузом с максимально возможной степенью эффективности.

Во многих вузах применяются различные информационные системы для автоматизации работы приемной комиссии [1]. Значительный объем информации, вводимый в процессе работы приемной комиссии, можно использовать для интеллектуального анализа данных с целью стратегического управления и планирования в рамках работы с контингентом абитуриентов и студентов.

Организация специального сбора информации для этих целей – трудоемкий и дорогостоящий процесс, поэтому для поддержки принятия решений представляется логичным использование данных приемных комиссий.

Проведение аналитической работы, по результатам работы приемной комиссии, предполагает, как правило, следующие задачи анализа данных:

- анализ количества поданных заявлений с различных школ, в которых была проведена профориентационная работа;
- анализ выбора направлений обучения и факультетов вуза абитуриентами различных школ, районов области;
- анализ качества подготовки выпускников различных школ, районов области;
- анализ географии абитуриентов вуза;
- анализ выбора вуза абитуриентами различных школ, районов области;
- анализ соотношения количества подавших документы и зачисленных в вуз, по различным школам, районам области.

- анализ количества абитуриентов по школам, районам области за различные годы (увеличение, уменьшение);
- анализ выбора абитуриентами отдельной школы факультетов и направлений обучения;
- анализ средних баллов, по результатам вступительных испытаний;
- анализ конкурса на отдельные факультеты (направления) за различные годы.

Эта информация необходима и вузу для определения перспектив развития и планирования профориентационной работы.

Встроенные в *Deductor* инструментальные средства для решения задач классификации, кластеризации, поиска ассоциативных правил, прогнозирования временных рядов, прогнозирования на основе линейных и нейросетевых моделей позволяют успешно решать указанные выше задачи анализа данных [2].

Встроенный в *Deductor* механизм визуализации обеспечивает удобное представление результатов анализа с точки зрения их интерпретации.

Важным преимуществом аналитической отчетности в *Deductor* является быстрый доступ к результатам анализа, не требующий от пользователя навыков анализа данных и работы в системе. При работе с отчетами пользователь не видит сценарий анализа данных, ему доступны только конечные результаты (выдержки) из работы аналитика.

Использование аналитической платформы *Deductor* способствует совершенствованию процедур оперативного анализа данных приемной комиссии и планированию стратегии профориентационной работы.

## **Литература**

1. Рыбанов А.А. Подходы к использованию информационных технологий в профориентационной работе. // Новые информационные технологии в образовании: матер. междунар. науч.-практ. конф. (Екатеринбург, 1-4 марта 2011 г.) / ФГАОУ ВПО "Рос. гос. профессионально-педагогический ун-т" [и др.]. – Екатеринбург, 2011. – Ч. I. – С. 234-237.
2. Рыбанов А.А. Информационные технологии в профориентационной работе // Школьные технологии. – 2011. – № 3. – С. 173-177.

## **НЕКОТОРЫЕ ПОДХОДЫ К ПРОГНОЗИРОВАНИЮ СПРОСА НА ПРОДУКЦИЮ ПРЕДПРИЯТИЯ С ИСПОЛЬЗОВАНИЕМ АП DEDUCTOR**

*Хомич С.В., ассистент,*

*Ровненский государственный гуманитарный университет, г. Ровно*

Как известно, прогнозирование продаж актуально практически для каждой компании. Качественный прогноз является первым шагом в решении множества бизнес-задач: оптимизация закупок, распределение ресурсов, минимизация кассовых разрывов, бюджетирование [1]. Для некоторых компаний прогнозирование становится жизненно важной задачей, которая может существенно способствовать финансовой стабильности и укреплению позиций на рынке.

Классический пример сказанного выше – ЗАО «Галичина», которая является одной из крупнейших компаний в Западной Украине по производству молочной продукции. Производство и реализация такого вида продукции имеет две главные особенности. Первая – это небольшой срок хранения молочной продукции, что определяет необходимость сопоставления скорости реализации (то есть объема реализации за определенный временной период) и складского запаса, который в свою очередь зависит от скорости производства. Вторая особенность обусловлена дефицитом молочного сырья на украинском рынке. Это в свою очередь ограничивает возможность быстрого наращивания производства при внезапном (вернее неожиданном) для компании и повышении спроса. Все это делает прогнозирование продаж (спроса) особенно важной задачей для оптимального планирования закупок сырья и производства продукции.

В данной статье предложены некоторые подходы к прогнозированию объемов реализации ЗАО «Галичина» как агрегированного показателя.

Первым и, наверное, самым важным условием качественного прогнозирования является наличие достаточного количества исходных данных. На данном этапе уже очень много предприятий Украины не только внедрили учетные системы класса ERP и CRM в свою деятельность, но и накопили в них достаточный для анализа объем данных (ЗАО «Галичина» не исключение). Поэтому проблема количества исходных данных перестает быть критической. Очевидно, что это не снимает требования к качеству данных.

Для решения задачи прогнозирования в контексте данной статьи был сформирован исходный набор данных, фрагмент которого представлен на рис. 1. В исходный набор данных попали ежемесячные данные о

реализации за период 2006-2009 г. плюс за девять месяцев 2010 года, измеренные денежных единицах, хотя также имеет смысл прогноз и в натуральных единицах.

	Date	Sales, ths UAH
▶	01.01.2006	7161,9
	01.02.2006	8790
	01.03.2006	10223
	01.04.2006	9353
	01.05.2006	9276,5
	01.06.2006	9954,4
	01.07.2006	9954,4

Рис. 1. Фрагмент исходных данных для прогноза общей реализации

Теперь более детально остановимся на сценариях прогнозирования. Было построено два типа моделей. Условно назовем их эндо- и экзо-моделями<sup>1</sup>.

Идея эндо-модели состоит в том, что объем реализации  $Y_t$  в месяце  $t$  зависит от объемов реализации в трех предыдущих месяцах:  $Y_t = F(Y_{t-1}, Y_{t-2}, Y_{t-3})$ . Данная идея предложена в работе [2, с. 678].

В случае с экзо-моделью сделаем предположение, что объем реализации в месяце  $t$  зависит от объема реализации в месяце  $t - 1$  ( $Y_{t-1}$ ), а также некоторых макроэкономических показателей: индекса инфляции месяца  $t - 1$  ( $I_{t-1}$ ), номинального обменного курса на конец месяца  $t - 1$  ( $ER_{t-1}$ ) и средней заработной платы по стране месяца  $t - 1$  ( $W_{t-1}$ )<sup>2</sup>.

$$Y_t = (Y_{t-1}, I_{t-1}, W_{t-1}, ER_{t-1}).$$

Данное предположение кажется справедливым, поскольку очевидно, что молочную продукцию покупают конечные потребители на часть дохода (зарплату), которая может обесцениваться через инфляцию. Если рост инфляции больше чем рост зарплаты, то падает покупательская способность и, соответственно, потенциальный объем реализации. Кроме этого, объем реализации может изменяться под воздействием обменного курса, который может влиять на объем экспорта и импорта.

<sup>1</sup> «Эндо-» и «экзо-» указывают на происхождение исходных переменных моделей по отношению к предприятию. Модель, исходные данные которой являются производственными показателями данного предприятия, имеют приставку «эндо-». Если же частично исходные данные модели являются внешними (например, макроэкономическими) данными, то она имеет приставку «экзо-».

<sup>2</sup> Значения показатели индекса инфляции и номинального обменного курсу доступны сразу после окончания месяца. Значение номинальной средней зарплаты публикуют к середине месяца, поэтому, в крайнем случае, можно взять за месяц  $t - 2$ .

Параметризация и получение прогнозных результатов для каждой из моделей сделана с помощью методов линейной регрессии и нейронной сети.

Перед построением моделей сделано следующее:

1. Произведена оценка исходных данных на предмет шумов, выбросов, пропусков. Последних обнаружено не было. Для сглаживания данных применена *Парциальная обработка*, результаты до и после которой показаны на рис. 2, а и 2, б.

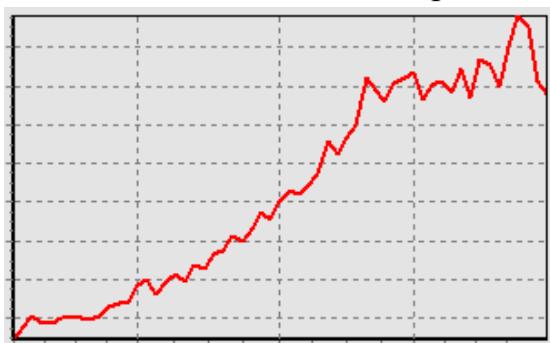


Рис. 2, а. До сглаживания данных

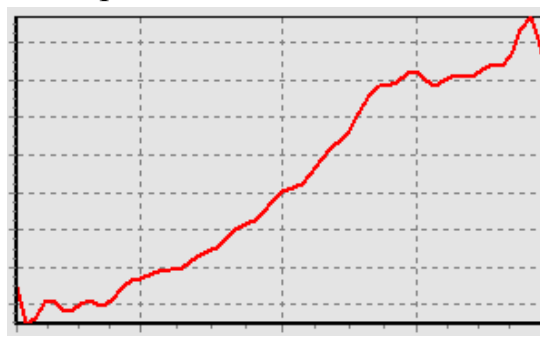


Рис. 2, б. После сглаживания данных

2. С помощью обработчика *Слияние с узлом* объединены данные о реализации и макроэкономические показатели.
3. С помощью обработчика *Скользящее окно* входные массивы данных преобразованы следующим образом:
  - для эндодеи объем реализации с глубиной погружения 3;
  - для экзодеи все 4 входных параметра с глубиной погружения 1.

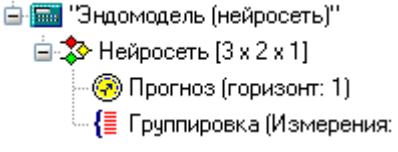
Ветки сценариев прогнозирования с использованием обработчиков *Линейная регрессия* и *Нейронная сеть* для экзо- и эндодеи ничем не отличаются и приведены в таблице 1 и таблице 2.

Таблица 1

Ветка сценария прогноза на основе линейной регрессии

Скриншот ветки	Описания узлов ветки
	<ol style="list-style-type: none"> <li>1. Делаем «метку» модели – <i>Калькулятор</i>.</li> <li>2. С помощью <i>Корреляционного анализа</i> выявляем факторы со степенью корреляции с выходным фактором меньше допустимой (0,2).</li> <li>3. Строим модель регрессии с помощью соответствующего обработчика.</li> <li>4. Делаем прогноз на 1 период вперед.</li> <li>5. Рассчитываем общую ошибку модели.</li> </ol>

## Ветка сценария прогноза модели на основе нейросети

Скриншот ветки	Описания узлов ветки
	<ol style="list-style-type: none"> <li>1. Делаем «метку» модели – <i>Калькулятор</i>.</li> <li>2. Строим нейросеть с с помощью соответственного обработчика. Структура сети: 3:2:1</li> <li>3. Делаем прогноз на 1 период вперед.</li> <li>4. Рассчитываем общую ошибку модели.</li> </ol>

Следует отметить, что после проведения корреляционного анализа были получены следующие результаты:

1. *Эндо-модель*: все входные параметры (реализация за три предыдущих месяца) сильно положительно коррелируют с выходным параметром (коэффициент – 0,93-0,99), что вполне очевидно и предсказуемо.
2. *Экзо-модель*: два входных параметра (средняя зарплата и обменный курс) имеет высокую степень положительной корреляции (>0,9), в то же время индекс инфляции – низкую (–0,13). Поэтому последний показатель из модели следует исключить. Данный результат (низкую корреляцию) можно объяснить вхождением молочной продукции в базовый потребительский набор, который слабо эластичен к цене.

Сам же процесс прогнозирования с использованием экзо- и эндо-подходов дал следующие результаты (таблица 3). Анализ общих ошибок модели позволяет сделать вывод о наилучшем качестве нейросетевой эндо-модели. По истечению прогнозируемого периода представилась возможность сопоставить реальные данные за октябрь 2010 года (37821 тыс. грн.) и данные прогноза за этот же период. В результате более точным оказался прогноз регрессионной эндо-модели.

Таблица 3

## Результаты прогнозирования

Метод	Вид модели	Ошибка	Прогноз на 01.10.10	Реальные данные за 01.10.10	Отклонение
<b>Эндо-модели</b>					
Регрессия	$Y_t = 631,24 + 1,75Y_{t-1} - 1,32Y_{t-2} + 0,56Y_{t-3}$	0,029	37243	37821	–578
Нейросеть	–	0,021	37067	37821	–754
<b>Экзо-модели</b>					
Регрессия	$Y_t = -8112,2 + 0,71Y_{t-1} - 7,43W_{t-1} + 921,08ER_{t-1}$	0,067	41079	37821	3258
Нейросеть	–	0,045	40661	37821	2840



Стоит отметить, что на основе эндомоделей можно делать прогноз с интервалом больше чем один период. В тоже время экзо модели следует использовать для прогноза на один период вперед, поскольку для прогноза на два и больше периода нет макроэкономических данных, хотя можно взять их прогнозное значение или сделать предположение об их неизменности.

### **Литература**

1. Прогнозирование продаж [Электронный ресурс]. – Режим доступа: <http://www.basegroup.ru/solutions/tasks/forecasting>.
2. Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных к знаниям (+ CD): Учеб. пособие. – 2-е изд., перераб. и доп. – СПб.: Питер, 2010. – 704 с.: ил.

## ПРОГНОЗИРОВАНИЕ РОССИЙСКОЙ СТАТИСТИКИ ПОЖАРОВ НА 2011 ГОД

*Татаринов В.В., доцент, Скударев М.А., студент,  
Московский Государственный Технический Университет  
имени Н. Э. Баумана, г. Москва*

Поставим задачу исследовать показатели пожарной обстановки РФ: абсолютное число пожаров и абсолютное число жертв, а также прямой материальный ущерб в год. Будем использовать статистические данные Министерства Российской Федерации по делам гражданской обороны, чрезвычайным ситуациям и ликвидации последствий стихийных бедствий (МЧС России).

Традиционный подход [1, 2] в анализе связи двух переменных  $y$  и  $x$

$$y = \hat{f}(x)$$

заключается в сопоставлении табличным данным парной регрессии. Далее проводится сопоставление этим данным линейной регрессии

$$y = a + bx + \varepsilon$$

или нелинейных регрессий.

Нелинейная регрессия относительно включенных в анализ объясняющих переменных, но линейных по оцениваемым параметрам:

- полиномы разных степеней, например,  $y = a + b_1x + b_2x^2 + b_3x^3 + \varepsilon$ ;
- равнобочная гиперболa  $y = a + b/x + \varepsilon$ .

Нелинейная регрессия по оцениваемым параметрам:

- степенная  $y = ax^b\varepsilon$ ;
- показательная  $y = ab^x\varepsilon$ ;
- экспоненциальная  $y = e^{a+bx}\varepsilon$ .

Далее проводится анализ тесноты связи изучаемых явлений через линейный коэффициент парной корреляции и индекс корреляции. Исследуется дисперсия зависимой переменной. Проводится  $F$ -тест на оценивание качества уравнения регрессии. Для оценки статистической значимости коэффициентов регрессии и корреляции рассчитываются  $t$ -критерий Стьюдента и доверительные интервалы каждого из показателей.

Не меньший, если не больший объем работы предстоит, если пытаться использовать модели временных рядов.

Таким образом, построение даже простейших моделей и анализ их адекватности достаточно сложен.

В тоже время, эти модели в том или ином виде реализованы в различных пакетах прикладных программ (например Deductor, Excel, Statgraphics, Statistica и др.). Их использование не требует высокой математической квалификации персонала. Особенно, если воспользоваться более широким пониманием методов Data mining.

В качестве возможного подхода, аналогичного задачам «обучение с учителем», проведём прогноз потерь от пожаров в Российской Федерации на 2011 год, исходя из официальной статистики пожаров МЧС 2003-2010 годы [3] (таблица 1).

Таблица 1

Год	Количество пожаров	Погибло человек	Ущерб, млрд. руб.
2003	239286	19257	4,8
2004	231486	18377	5,8
2005	226952	18194	6,8
2006	218570	17065	7,9
2007	211163	15924	8,6
2008	200386	15165	9,8
2009	187490	13933	10,9
2010	179098	12983	11,70
2010	175000	12900	11,76
2011	161000	11700	12,76

С помощью пакета прикладных программ Excel подберём лучший прогноз из данных 2003-2009 гг. на 2010г. и выберем лучший с учётом реальных данных на этот год. Лучшей моделью оказалась полиномиальная регрессия второй степени. Погрешность прогноза на 2010 год не превышает 2,5 %.

В выбранную модель подставим данные за 2010 г. и получим прогноз на 2011 год. Официальный прогноз [3]: «в 2011 году **количество погибших при техногенных пожарах и на водных объектах, взрывов бытового газа** прогнозируется на уровне 2010 года (не более: 13 000 погибших при техногенных пожарах)...».

В построенной модели этот показатель имеет следующий вид (рис. 1-3).



Рис. 1. Количество пожаров

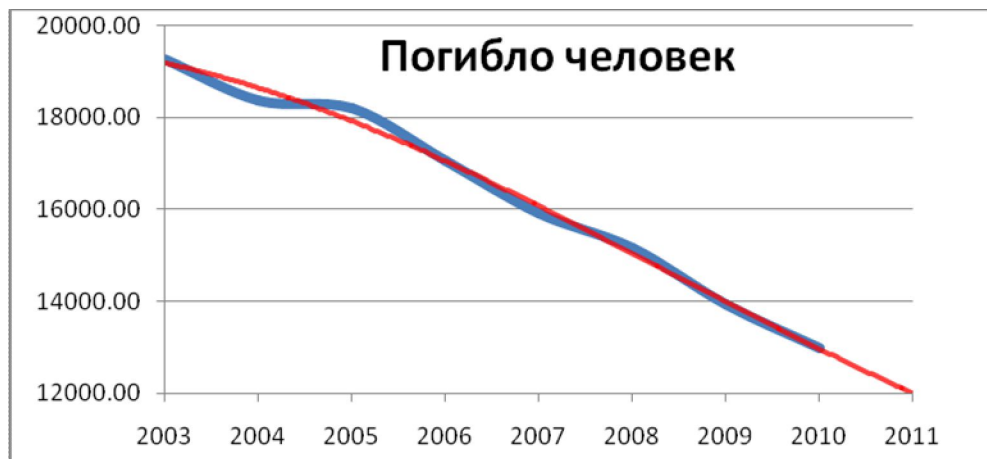


Рис. 2. Количество погибших на пожарах



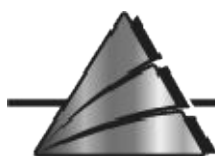
Рис. 3. Ущерб от пожаров

По прогнозу полученной модели в 2011 году произойдет 161 тыс. пожаров, что на 10,1% меньше чем в 2010 году. Количество погибших составит не менее 11 тыс. 700 человек (уменьшение по сравнению с 2010

годом составит 9,8%). Прямой материальный ущерб составит 12,76 млрд. руб., что на 8,5% больше по сравнению с 2010 годом.

### **Литература**

1. Четыркин Е.М. Статистические методы прогнозирования. Изд. 2-е, перераб. и доп. – М.: Статистика, 1977.
2. Эконометрика: Учебник // Под ред. И.И. Елисейевой – М.: Финансы и статистика, 2003.
3. Официальный сайт МЧС России <http://www.mchs.gov.ru>.



# BaseGroup Labs

ТЕХНОЛОГИИ АНАЛИЗА ДАННЫХ

**BaseGroup Labs** – профессиональный поставщик продуктов и решений в области анализа данных. Мы имеем многолетний опыт работы в области разработки аналитических алгоритмов и создания законченных систем. BaseGroup Labs предлагает полностью интегрированные продукты, объединяющие все необходимые инструменты анализа: хранилища данных, аналитическую отчетность, механизмы поиска закономерностей и построения моделей, средства интеграции аналитических систем с платформами сторонних производителей.

Системы от BaseGroup Labs выполнены с применением **самых современных** информационных технологий.

## Технологии

### Data Warehouse

хранилище данных

- Консолидация анализируемых данных, обеспечение непротиворечивости данных
- Быстрый доступ к необходимой информации
- Автоматическое обновление данных
- Богатый семантический слой

### Data Mining

добыча данных

- Прогнозирование
- Поиск закономерностей и зависимостей
- Извлечение правил
- Оптимизация процессов
- Анализ по принципу «что-если»

### OLAP

многомерный анализ данных

- Многомерная отчетность, позволяющая извлечь максимум полезной информации из имеющихся данных
- Гибкие механизмы навигации и манипулирования данными
- Анализ тенденций
- Простота использования конечным пользователем

### Knowledge Discovery in DB

обнаружение знаний в базах данных

- Механизмы улучшения качества исходных данных (очистка, преобразование и трансформация данных)
- Построение сценариев обработки данных
- Механизмы построения моделей
- Интеграция моделей в информационные системы

Наши системы базируются на **собственном аналитическом ядре**, что обеспечивает беспрецедентную **гибкость** при выборе способов анализа и создании прикладных решений. Применение самообучающихся механизмов дает возможность быстрой адаптации решения под постоянно изменяющиеся условия.

Россия, 390046, г. Рязань, Введенская, д. 115, оф. 447

Т./ф.: +7 (4912) 24-09-77, +7 (4912) 24-06-99

[info@basegroup.ru](mailto:info@basegroup.ru)

[www.basegroup.ru](http://www.basegroup.ru)



Deductor – флагманский продукт BaseGroup Labs, концентрирующий многолетний опыт компании и вобравший в себя самые удачные архитектурные идеи и современный математический аппарат. В Deductor реализованы технологии анализа структурированных данных: нейронные сети, деревья решений, хранилища данных и OLAP, ассоциативные правила, карты Кохонена и многое другое. Использование Deductor в учебном процессе поможет студентам освоить алгоритмы машинного обучения и системы интеллектуальной обработки информации на практике, решая актуальные задачи по консолидации, очистке, прогнозированию, классификации, кластеризации, скорингу.

## Образование

Для высших учебных заведений BaseGroup Labs предлагает специальные условия. Заключив с нами соглашение о сотрудничестве, преподаватели и сотрудники учебного заведения получают следующие возможности:

- Аналитическую платформу **Deductor Academic** для проведения практикумов по дисциплинам, связанным с информационно-аналитическими системами, интеллектуальными информационными системами, системами поддержки принятия решений и другим курсам для прикладных информатиков и экономистов.
- Бесплатное e-learning обучение преподавателей на образовательном портале **edu.basegroup.ru** в полноценной системе **дистанционного обучения** и сертификацию по результатам обучения, обсуждение возникающих вопросов на форуме.
- Большое число **методических разработок** для проведения практических занятий со студентами по всем современным технологиям анализа данных.

Участие в программе полностью **бесплатное**. Образовательная инициатива действует с 2005 года и за это время более **90** вузов России, Украины и Беларуси стали нашими партнерами и используют аналитическую платформу Deductor в учебном процессе. Вот некоторые из них:

- Российская экономическая академия имени Г.В. Плеханова;
- Государственный университет управления;
- Московский авиационный институт;
- Санкт-Петербургский государственный университет;
- Белорусский государственный университет информатики и радиоэлектроники.

Форму и условия партнерства, полный список вузов-партнеров и другую дополнительную информацию можно получить на образовательном портале <http://edu.basegroup.ru>.

Россия, 390046, г. Рязань, Введенская, д. 115, оф. 447

Т./ф.: +7 (4912) 24-09-77

+7 (4912) 24-06-99

education@basegroup.ru

**БИЗНЕС-АНАЛИТИКА. ИСПОЛЬЗОВАНИЕ АНАЛИТИЧЕСКОЙ  
ПЛАТФОРМЫ DEDUCTOR В УЧЕБНОМ ПРОЦЕССЕ ВУЗА:**

Сборник материалов  
II Межвузовской научно-практической конференции

28 июня 2011 года