

# Программная поддержка интеллектуального анализа данных

**Ветров Д. Е.**

Международный научно-исследовательский  
институт проблем управления  
пр-т 60-летия октября, 9, г.Москва,  
Россия  
*vetrov@irias.ru*

**Чернова С. С.**

Институт системного анализа РАН  
пр-т 60-летия октября, 9, г.Москва,  
Россия  
*chernova@cpt-ran.ru*

## Аннотация

В докладе описывается способ построения и пример реализации интегратора программных модулей, обеспечивающего построение различных схем обработки данных на базе автоматически сгенерированных процедур обработки, существенно зависящих от данных.

## 1 Введение

В процессе проектирования и создания сложного технического объекта рассматриваются и сравниваются различные технические решения, касающиеся структуры объекта, механизмов его функционирования, выбора параметров и других элементов объекта.

Значительный прогресс в области математического моделирования и возможностей вычислительной техники позволяет провести широкий спектр исследований без проведения натуральных экспериментов. Этот факт делает математическое моделирование (вычислительные эксперименты для исследования аналитических моделей создаваемого объекта и окружающей его среды) одним из самых распространенных методов анализа и оптимизации структуры технических объектов.

Для сокращения времени проектирования и числа дорогостоящих натуральных экспериментов, а также для улучшения качества проектирования создаются компьютерные системы проектирования для поддержки принятия инженерных решений. Такие системы, являющиеся, по существу, специализированными системами моделирования и поддержки принятия решений, позволяют исследовать большое количество вариантов построения объекта (конфигурации, параметров и др.), предсказывать ожидаемые характеристики и находить наилучшие (рациональные) решения.

Современные когнитивные технологии открывают новые возможности для систем моделирования и принятия инженерных решений. В первую очередь эти возможности связаны с технологиями построения вычислительных моделей, основанных на данных. Особенность таких моделей состоит в том, что они «обучаются» на некотором множестве данных-прототипов и гарантированно могут использоваться только для тех входных данных, которые подобны входным данным прототипов, с помощью которых была построена модель. Другими словами модели должны перестраиваться при смене обучающего множества данных.

Эти возможности в свою очередь ставят новые задачи для программной поддержки систем моделирования и анализа данных:

- автоматическая генерация процедур обработки данных;
- интеграция существующих и сгенерированных процедур в единую компьютерную модель вычислений.

В докладе описывается способ построения и пример реализации интегратора программных модулей, обеспечивающего построение различных схем обработки данных на базе автоматически сгенерированных процедур обработки, существенно зависящих от данных.

## 2 Задачи программного интегратора

**Интегратор программного обеспечения** представляет собой инструментальное средство, предназначенное для разработки и прогона вычислительных схем, которые создает пользователь в

диалоговом режиме из отдельных компьютерных моделей обработки данных.

**Компьютерная модель** - это программный компонент, реализующий некоторый математический алгоритм обработки данных или его часть. Как элемент программного обеспечения компьютерная модель реализуется на основе динамически загружаемых библиотек. Таким образом, функциональность модулей компьютерных моделей подключается к интегратору на стадии его выполнения, а не на стадии сборки (компиляции).

**Вычислительная схема** представляет собой ориентированный граф, в котором узлы соответствуют отдельным компьютерным моделям, а дуги последовательность взаимодействия этих моделей.

При построении вычислительной схемы компьютерные модели могут выбираться либо из множества уже созданных моделей, либо создаваться заново. Для того чтобы создавать новые компьютерные модели в процессе проектирования вычислительной схемы, интегратор программного обеспечения должен иметь возможность подключения автоматических генераторов процедур.

Таким образом, интегратор программного обеспечения объединяет в единую систему обработки и анализа данных не только отдельные процедуры обработки данных, но и генераторы, с помощью которых эти процедуры были получены. Причем, как генераторы процедур, так и созданные ими процедуры являются по отношению к интегратору внешними программными компонентами. Они могут использоваться независимо при разработке других инженерных приложений. По отношению к этим программным моделям интегратор представляет собой лишь графическую оболочку, к которой подключаются программные модели. Причем, активация программных моделей происходит на этапе вызова их функций, а не на этапе сборки интегратора.

Таким образом, интегратор программного обеспечения предоставляет пользователю графический интерфейс для решения следующих задач:

- формирование различных вычислительных схем путем комбинирования различных компьютерных моделей из заданного множества;
- задание входных данных для проведения экспериментов и проведение экспериментов;
- сохранение и просмотр процессов (прохождение и завершение этапов) и результатов экспериментов;
- отладка и отработка логики вычислений и форматов взаимодействия отдельных компьютерных моделей;
- пополнение множества компьютерных моделей путем генерации новых процедур, реализующих зависящие от данных алгоритмы обработки, без компиляции и сборки программного обеспечения интегратора.

### **3 Пример построения интегратора программного обеспечения**

#### **3.1 Цели разработки**

Авторами был разработан интегратор программного обеспечения, который представляет собой Windows приложение, имеющее графический пользовательский интерфейс и позволяющее

- создавать основанные на данных программные модули по заданному множеству данных с использованием автоматических генераторов процедур;
- проводить валидацию созданных сгенерированных процедур путем их применения к заданному тестовому множеству данных;
- формировать схемы вычислительных процессов с участием созданных процедур;
- проводить вычислительные эксперименты.

В разработанном интеграторе использовались следующие автоматические генераторы процедур:

- генератор процедур снижения размерности;

- генератор процедур аппроксимации многомерных данных;
- генератор процедур генерации геометрии объекта.,

Интегратор предназначен для инженеров и аналитиков в области обработки данных. Созданные основанные на данных программные модули предназначены для использования разработчиками программного обеспечения для разработки приложений:

- с использованием C++ компилятора;
- в среде Python.

### 3.2 Функциональность интегратора

Разработанный интегратор обеспечивает следующую функциональность:

- загрузку входного (пользовательского) множества данных в виде файла (файлов) заданных форматов;
- подключение произвольной пользовательской программы, реализованной в виде Модуля импорта, и проведение вычислительных экспериментов с Модулем импорта для получения входного множества данных;
- автоматическую генерацию программной модели по заданным данным с использованием выбранного генератора процедур;
- валидацию созданной программной модели путем применения его к заданному пользователем множеству тестовых данных;
- регистрацию программной модели для ее дальнейшего совместного функционирования с интегратором;
- проектирование, валидацию и прогон схемы вычислительного процесса.

Дополнительно в состав интегратора включены следующие модули:

- вычисления набора стандартных характеристик описательной статистики, традиционно используемого в математических пакетах;
- модуль сравнения качества сгенерированных процедур одного класса;
- модуль управления логированием, позволяющий перенаправлять в указанные пользователем файлы потоки сообщений о регистрации всех событий, возникающих в процессе прогона вычислительной схемы и представляющих интерес для пользователя (ошибки, предупреждения и другая информация).

## 4 Применение программного интегратора

Разработанный интегратор программного обеспечения был успешно применен для решения задачи генерации аэродинамических профилей крыла.

Задача генерации профилей крыла решалась следующим образом. Поскольку цифровое описание исходного объекта представляет собой вектор большой размерности, то вначале необходимо было найти многообразие меньшей размерности с достаточной точностью приближающее исходное множество данных. Генерация нового объекта проводится в пространстве малой размерности, после чего восстанавливается полноразмерное описание вектора.

Для решения этой задачи с помощью интегратора программного обеспечения была построена следующая вычислительная схема:

- задается произвольное множество, состоящее из описаний профилей гражданских пассажирских самолетов, и требование к точности создаваемой процедуры снижения размерности;

- с помощью генератора процедур автоматически создается процедура снижения размерности;
- выбирается произвольное подмножество заданного множества профилей, которое будет входными данными (множеством прототипов) для процедуры автоматического создания процедур случайной генерации;
- с помощью генератора процедур автоматически создается процедура случайной генерации профилей;
- генерируются новые профили с помощью автоматически полученной процедуры случайной генерации.

Применение этой схемы к разным наборам исходных данных (прототипам) и проведение вычислительных экспериментов с автоматически построенными процедурами снижения размерности и генерации геометрии показали, что с помощью такой схемы можно генерировать новые аэродинамические профили с заданными или прогнозируемыми свойствами.

## Список литературы

- [1] Кулешов А.П. Когнитивные технологии в адаптивных моделях сложных объектов. *Информационные технологии и вычислительные системы*, в. 1, 2008, с. 18 – 29.
- [2] Бернштейн А.В., Кулешов А.П. Математические методы в когнитивных инженерных технологиях. Обзорные прикладной и промышленной математики, сер. *Вероятность и статистика*. 2008, т. 15, № 3, с. 451 – 452
- [3] Бернштейн А.В., Бурнаев Е.В., Дорофеев Е.А., Свириденко Ю.Н., Чернова С.С. О решении некоторых задач анализа данных, возникающих при построении адаптивных суррогатных моделей сложных объектов. Научно-теоретический международный журнал *Искусственный интеллект*, Донецк, 2008, вып. 4, с. 40-48.
- [4] Бернштейн А.В., Кулешов А.П. Когнитивные технологии в проблеме снижения размерности описания геометрических объектов. *Информационные технологии и вычислительные системы*. 2008, №2, с. 6 – 19.
- [5] Иванова Е.П., Чернова С.С. Снижение размерности сложных геометрических объектов при наличии частных параметрических моделей. *Информационные технологии и вычислительные системы*, 2008, №4.
- [6] Бернштейн А.В., Бурнаев Е.В., Дорофеев Е.А., Свириденко Ю.Н., Чернова С.С. Каскадные процедуры снижения размерности. *Труды Одиннадцатой национальной конференции по искусственному интеллекту с международным участием* (Дубна, 2008), т. 1, с. 241 – 250.