

ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ АСНИ НА ОСНОВЕ  
МЕТОДОВ АВТОМАТИЧЕСКОЙ КЛАССИФИКАЦИИ

(г. Томск)

В настоящее время получили распространение технические системы, в которых существенную роль играют компоненты природной среды. Проектирование таких систем требует проведения трудоемких натурных экспериментов. С целью сокращения затрат на проведение экспериментов необходимо использовать системы автоматизации натуральных исследований. Однако применение систем автоматизации экспериментов, традиционно ориентированных на автоматизацию процессов сбора, измерения, предварительной обработки, учета систематических искажений, анализа и интерпретации результатов наблюдений, не позволяют кардинально сократить сроки и затраты на проведение прикладных натуральных экспериментов в различных областях науки ввиду невозможности активного экспериментирования с факторами воздействия компонентов природной среды на исследуемые объекты. Одной из особенностей названных натуральных экспериментов является чрезвычайно большой объем данных, поэтому одной из важнейших задач построения систем автоматизации научных исследований является решение задач сжатия информации.

Цель данной работы - изложение вопросов информационного и программного обеспечения АСНИ, основанных на применении методов автоматической классификации для сжатия данных и имитационного моделирования, методология применения которого изложена в работе [1], для восстановления данных.

Исходную информацию об исследуемом объекте с природными компонентами представим в виде таблицы типа "объект-свойство", где под  $i$  -м объектом будем понимать многомерное измерение в  $i$  -й момент времени, а под свойством - значения характеристик объекта в  $i$  -й момент времени. Тогда информацию об объекте можно хранить в ЭВМ в виде матрицы, в которой строки соответствуют объектам, а столбцы - свойствам объектов. Размеры матриц, получаемых в прикладных натуральных экспериментах, велики, и для размещения их в памяти ЭВМ требуется несколько томов внешних запоминающих устройств, поэтому поиск информации в таких таблицах является практически неразрешимой задачей.

В целях более эффективного использования больших объемов информации для принятия решений, ее приходится агрегировать, объединять по каким-либо критериям в группы с последующим усреднением по группе, классифицировать, иными словами, сжимать информацию с минимальными потерями перед вводом ее в хранилище. Наиболее пригодными для выполнения процедуры сжатия информации являются методы автоматической классификации, обзор которых дан в [2]. При этом всю информацию об исследуемых объектах будем рассматривать как совокупность элементов, являющихся многомерными (вектор  $\vec{S}$ ). Требуется разбить данную совокупность на такие группы, для каждой из которых можно подобрать одну адекватную имитационную модель, входами которой являются параметры модели (вектор  $\vec{P}$ ), а выходами - набор свойств исследуемых объектов (вектор  $\vec{S}$ ). Информация о таких классах может быть введена в ЭВМ в виде таблицы типа "объект-свойство", где под объектом будем понимать кластеры, а под свойствами - характеристики эталонов кластеров.

Процедуру создания подобной таблицы можно рассматривать как последовательность следующих шагов:

1. Автоматическая классификация множества элементов на кластеры.

2. Определение эталонов кластеров.

3. Идентификация векторов параметров имитационной модели по векторам характеристик эталонов каждого класса.

4. Проверка адекватности модельной и экспериментальной выборок каждого класса.

5. В случае отклонения гипотезы адекватности - повторное разбиение класса и повторение шагов 2 - 4.

Данный подход к задаче сжатия информации позволяет с небольшими временными и вычислительными затратами создавать матрицу, компактно описывающую исходную совокупность данных, занимающую небольшой объем памяти ЭВМ. Структура матрицы, хранящей сжатые данные, приведена на рис.1. Размеры матрицы невелики, что позволяет при работе с ней обходиться только быстрой памятью ЭВМ - это ускоряет процесс поиска полезной информации.

Для увеличения достоверности информации, получаемой по матрице сжатых данных, предлагается провести модификацию последней, разбивая каждый класс на подклассы. Для этого разработан алгоритм автоматической классификации, основанный на вариационном подходе и производящий разбиение на подклассы в зависимости от значения критериальной функции, аргументом которой являются потери инфор-

N класса	Мощность класса	$\vec{P}$	$\vec{S}$	Счетчик пополнений класса	Номера точек в классе

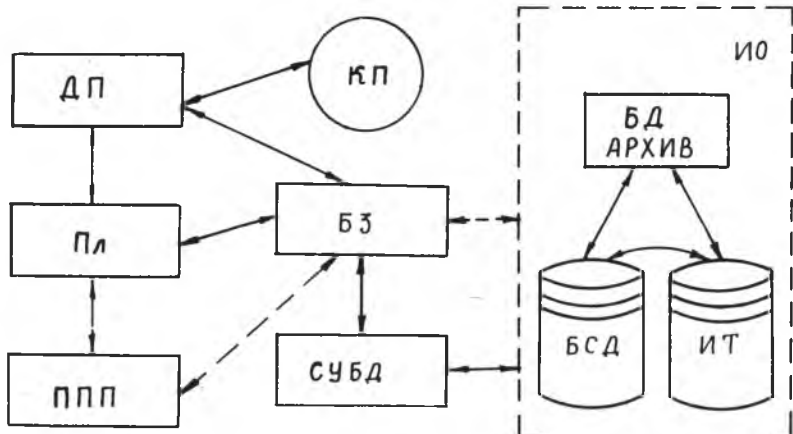
Р и с. 1. Структура матрицы, хранящей сжатые данные

матрицы, которые возникают при выборе количества и размеров классов. Алгоритм заключается в дихотомическом делении точек каждого класса с проверкой в каждом получаемом подклассе критерия качества. Если разбиение удовлетворяет заданной точности, то алгоритм заканчивает свою работу, в противном случае производится дихотомическое деление одного из полученных подмножеств. Если потери количества информации находятся в требуемых пределах, то данное подмножество называется подклассом. Точки, вошедшие в данный подкласс, исключаются из исходного множества и алгоритм повторяется. В результате работы данного алгоритма все точки класса будут разбиты на подклассы так, что суммарная потеря количества информации не превысит заранее заданной величины.

В каждом подклассе выбирается точка - представитель класса. Все точки, вошедшие в данный подкласс, проверяются на роль представителя. Из них выбирается та, относительно которой дисперсия в данной группе точек будет минимальна.

Для автоматизации процедуры сжатия информации описанным способом и для возможности работы пользователей - непрограммистов с данными, обработанными подобным образом, предлагается программно-инструментальный комплекс, общая структура которого приведена на рис.2.

Пользователь системы (ПС) общается с системой на ограниченном естественном языке посредством диалогового процессора (ДП). Проанализировав запрос, планировщик системы (Пл) формирует алгоритм, в соответствии с которым организуется последовательность выполнения программных модулей, выбранных из пакета прикладных программ (ППП). Данные и программы описаны в базе знаний (БЗ). В данной системе



Р и с. 2. Общая структура программно-инструментального комплекса

информационное обеспечение (ИО) представлено совокупностью баз данных "Архив", в которой хранятся результаты пассивных наблюдений за системой с природными компонентами, баз сжатых данных (БСД) и таблицей, хранящей информацию о классах и характеристиках их представителей.

- В системе предусмотрено три режима работы:
- создание базы сжатых данных;
  - пополнение базы сжатых данных;
  - поиск информации в базе сжатых данных.

Процедура "Создание БД" предусматривает создание в памяти ЭВМ двух наборов данных: базы сжатых данных и таблицы с перечнем классов и их характеристик. В результате работы процедуры формируется файл с числом записей, равным количеству классов, каждая из которых содержит следующие поля: номер класса, мощность класса, параметры имитационной модели, вектор характеристик объекта - представителя данного класса, значение счетчика числа пополнений класса и номера объектов, вошедших в данный класс.

Процедура "Пополнение БД" выполняется при созданной таблице описания класса, если в процессе исследования появились новые данные натуральных экспериментов. Отнесение каждого нового данного к классу выполняется методом распознавания образов. Если класс не

обнаружен, то создается новый класс, для которого автоматически рассчитываются все необходимые атрибуты. Соответствующая информация пополняет таблицу сжатых данных. Если количество вновь вводимых данных больше порогового числа, то производится их предварительная классификация с выделением в каждой полученной группе представителя (эталона). Процедура распознавания образов применяется только к представителям групп с последующим отнесением всех точек группы в определенный класс. Если класс, соответствующий искомой точке, уже существовал и описан, то счетчик пополнения классов фиксирует это событие и увеличивает свое содержимое на единицу (при отнесении к классу сразу  $N$  точек содержимое счетчика увеличивается на  $N$  единиц). По достижении счетчиком порогового числа все характеристики данного класса уточняются.

Процедура "Поиск в БД" позволяет по запросу пользователя выдать информацию с заданной точностью. Если точность не устраивает пользователя, то, воспользовавшись значениями вектора параметров имитационной модели  $\vec{P}$ , можно восстановить данные, "проиграв" на ЭВМ поведение исследуемой системы с природными компонентами.

Информационное обеспечение, как следует из изложенного, может быть реализовано на основе стандартных систем управления базами данных (СУБД). Программное обеспечение в связи с необходимостью создания расширяемой системы реализовано на основе модульного принципа, что позволяет использовать стандартные пакеты прикладных программ. В настоящей работе программное обеспечение включает модули имитационного моделирования, реализованные с использованием ППП САС [3], а также модули автоматической классификации и распознавания образов, описанные средствами ППП ОТЭКС [4]. Управляющая программа реализована на универсальном языке ПЛ/1, что позволяет проводить ее стыковку со стандартными СУБД и организовать диалоговый режим работы.

#### Библиографический список

1. Полищук Ю.М. Особенности построения моделей сложных систем с природными компонентами // Применение методов теории информации для повышения эффективности и качества сложных радиоэлектронных систем. - М.: Радио и связь, 1984. С. 14-16.
2. Дорофеев А.А. Алгоритмы автоматической классификации // Автоматика и телемеханика, 1971. № 12. С. 78-113.

3. Варламов Г.В. и др. Имитационное моделирование технических систем оперативного обнаружения аномалий на поверхности океана //Мониторинг океана.-Л.:Ленингр.университет, 1986. С.40-53.

4. Загоруйко Н.Г., Лбов Г.С., Машаров Ю.П. Пакет прикладных программ для обработки таблиц экспериментальных данных ОТЭКС// Вопросы кибернетики. М., 1977. Ч.2. С.5-9.

УДК 681.3.06

С.А.Фоменков, В.А.Гришин, С.Г.Колесников

АВТОМАТИЗИРОВАННАЯ ИНФОРМАЦИОННО-ПОИСКОВАЯ  
СИСТЕМА ПО ФИЗИКО-ТЕХНИЧЕСКИМ ЭФФЕКТАМ

(г. Волгоград)

В рамках создания автоматизированного банка данных по физико-техническим эффектам и конструкторско-технологическим решениям разработана автоматизированная информационно-поисковая система (АИПС) по физико-техническим эффектам (ФТЭ). Система предназначена для организации автоматизированного многоаспектного поиска физико-технических эффектов, выдачи пользователю фактографической информации о них; централизованного сбора и накопления сведений о ФТЭ.

Режим автоматизированного поиска, реализуемый АИПС, позволяет значительно сократить время получения необходимых сведений, устранить необходимость рутинной переработки и систематизации физических знаний, содержащихся в различных печатных источниках информации (справочниках, учебниках, монографиях, научных статьях и т.п.). Благодаря этому существенно повышается объем активно используемых знаний по физике при выполнении различных научно-исследовательских и проектно-конструкторских разработок, а также в учебном процессе.

Разработанная система оперирует структурированными физическими знаниями в виде так называемых физико-технических эффектов. Каждый ФТЭ представляется трехкомпонентной структурой [1-2]:

$A \rightarrow B \rightarrow C,$

где А - входное воздействие (вход),

В - объект (материалоноситель),

С - выходное воздействие (выход).