

5. Ефимов А. Н., Криворуков Е. В. О выборе рациональной последовательности оптимальных линейных статистических операций и функциональных преобразований случайных процессов.— Автоматика и телемеханика, 1977, № 7.
6. Ефимов А. Н., Криворуков Е. В. Некоторые особенности усреднения малых выборок при косвенных наблюдениях.— В кн.: Вопросы атомной науки и техники. Сер. Автоматизация и математическое обеспечение физического эксперимента. Харьков, 1977, вып. 1 (7).
7. Кульман Н. К., Титаренко Н. В. Интегрирование непрерывного марковского процесса в случае косвенного наблюдения.— Автометрия, 1977, № 1.
8. Феллер В. Введение в теорию вероятности и ее приложения.— М.: Мир, 1967, т. 2.
9. Кендалл М. Дж., Стьюарт А. Теория распределений.— М.: Наука, 1966.
10. Корн Г., Корн Т. Справочник по математике.— М.: Физматгиз, 1974.
11. Grow E. L., Siddiqvi M. M. Robust Estimation of Location.— J. American Statistical Association, 1967, vol. 62, N 318, p. 353—389.
12. Виленкин С. Я. Статистическая обработка результатов исследования случайных функций.— М.: Энергия, 1979.

*Поступила в редакцию 26 марта 1981 г.;
окончательный вариант — 15 июля 1982 г.*

УДК 681.3.068

Ю. А. ГОРИЦКИЙ, С. Е. ЖАРИНОВ
(Петропавловск-Камчатский)

ОРГАНИЗАЦИЯ ГРАФИЧЕСКОЙ ДИАЛОГОВОЙ СИСТЕМЫ АНАЛИЗА ДАННЫХ

Введение. Задачи анализа многомерных данных, где признаки носят количественный характер и их число невелико, типичны для естественных исследований [1]. В настоящее время их решение зачастую связывают с диалоговым подходом [2], когда главная роль в анализе (визуальном) структуры и оценке результатов отводится самому исследователю, а ЭВМ выполняет всю вспомогательную работу по преобразованию информации к виду, удобному для наглядного представления. Такой подход в различной степени реализован в ряде систем распознавания образов, классификации и отображения информации [3—7]. Однако, как указано в [2], большей частью они не отвечают требованиям, которым должна удовлетворять диалоговая система анализа данных, и в первую очередь требованиям по организации, поскольку основу их составляют заранее определенные наборы алгоритмов со своими специфическими взаимосвязями.

В качестве альтернативы может быть предложена предварительная разработка вопросов организации с учетом всех аспектов задачи анализа данных, в том числе получение соответствующего «скелета» системы. В работе приводится описание структурной схемы, организации данных и диалога системы «Вулканит» Института вулканологии ДВНЦ АН СССР [8]. Первый вариант системы реализован на ЭВМ «Минск-32», второй — на ЭВМ ЕС-1033, однако рассматриваемые вопросы являются машинно-независимыми.

Структура системы. Состав системы и ее структура определяются классом решаемых задач. Как известно, процесс анализа данных обычно носит итерационный характер, где каждая итерация состоит из трех шагов: выбор данных для анализа, выбор способа анализа из имеющегося набора и формы представления результатов, оценка результатов.

Первые два выполняются совместно исследователем и ЭВМ, поэтому система должна содержать возможности манипуляций над данными, графического отображения и собственно способы анализа, которые условно можно разделить на три группы в зависимости от имеющейся априорной информации. Так, при наличии модели изучаемой структуры (например, статистической) интерес представляет определение параметров этой мо-

дели. Если есть лишь предположение о том или ином виде структуры, то возникает задача проверки соответствующей гипотезы. Однако самой распространенной является ситуация, когда априорная информация отсутствует. Единственным способом поиска закономерностей тогда может быть метод «проб и ошибок», сочетающий опыт и интуицию специалиста-предметника с возможностями ЭВМ по обработке, хранению и представлению данных. Наибольшее значение в этом случае приобретает снижение размерности исходного пространства с последующим отображением исследуемой совокупности в пространстве меньшей размерности.

Таким образом, очевидна важность наличия в системе указанных методов и моделей анализа данных. С другой стороны, опыт применения систем PROMENADE [9] и OLPARS [10] показывает, что даже небольшой набор таких средств позволяет успешно решать широкий круг практических задач за счет гибкой структурной организации. Поэтому для увеличения функциональных возможностей системы главную роль играют вопросы взаимосвязи отдельных групп алгоритмов (блоков).

При разработке системы «Вулканит» было испробовано несколько вариантов структурной организации: линейный, циклический, древовидный. Однако наиболее удобной оказалась комбинированная схема, представленная на рис. 1. Опишем коротко назначение отдельных блоков и связи между ними.

Блок предварительной обработки служит для подготовки исходных данных к проведению очередного этапа исследования путем выделения из них нужных в данный момент признаков и объектов, усреднения объектов, а также начального образования классов. На выходе блока формируются рабочие данные.

Блоки статистического анализа, проверки гипотез и преобразования соответствуют упомянутым выше трем группам способов. В реализованном варианте системы «Вулканит» первый из них содержит модели корреляционного, регрессионного анализа и анализа нормальных смесей; второй — статистические тесты проверки гипотез о нормальности, уни-

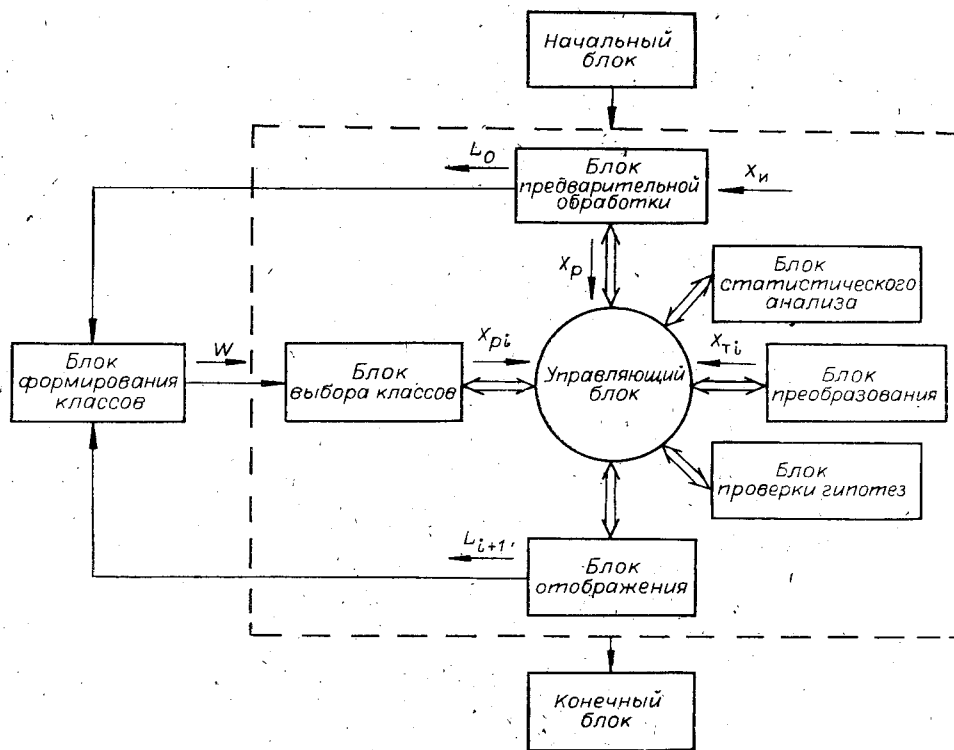


Рис. 1.

модальности и однородности; третий — методы перехода к новым координатам — главным компонентам, оптимальным дискриминантным векторам, а также признакам, задаваемым в явном виде самим исследователем из класса дробно-линейных функций. Образованные различными методами новые признаки приформировываются друг к другу, составляя преобразованные данные. Указанный набор возможностей, естественно, не претендует на полноту, но при необходимости может быть расширен, например, за счет алгоритмов кластер-анализа, нелинейного шкалирования и т. д.

В блоке отображения сосредоточены программы наглядного графического представления результатов в виде одномерных и двумерных гистограмм, диаграмм разброса, разнообразных сглаженных картин, а также средства выделения увиденных закономерностей непосредственно по изображениям.

Блок формирования служит для образования и корректировки информации о классах, заданных на этапе предварительной обработки либо в процессе анализа графического материала. Эта информация используется в блоке выбора, который реализует возможность изучения структуры данных по частям путем выделения из анализируемой совокупности одного или нескольких классов для отдельного исследования (при этом образуются соответствующие классифицированные данные). Последний блок содержит, кроме того, операций пересечения, объединения, разности, исключения, а также распечатки содержимого классов.

Блок управления организует переходы от одной группы возможностей системы к другой, выбор необходимых данных и способов их обработки, предоставляя тем самым исследователю самому определять и осуществлять ту или иную стратегию анализа.

Начальный и конечный блоки выполняют вспомогательные функции, связанные с обработкой служебной информации, характеризующей состояние системы к моменту начала или конца работы. Особое значение эти функции приобретают в том случае, когда в процессе диалога имело место прерывание, вызванное самим пользователем, и необходимо сохранить все полученные к этому времени результаты до следующего сеанса исследования.

Организация данных. Вопрос организации данных имеет два аспекта. Первый — работа с классами.

Обычно в диалоговых системах принята иерархическая схема [4], когда исходная совокупность представляет собой объединение непересекающихся классов первого уровня, те, в свою очередь, могут распасться на непересекающиеся группы классов второго уровня и т. д. Недостатком ее является, например, невозможность разбить одно и то же множество данных на подмножества несколькими способами с сохранением результатов всех делений. Этого, а также некоторых дополнительных удобств можно добиться, допустив задание пересекающихся классов (рис. 2).

В системе «Вулканит» для работы с классифицированными данными вводится понятие характеристического вектора. Его длина равна общему числу исследуемых объектов, а каждый элемент содержит информацию о принадлежности соответствующего объекта одному или одновременно нескольким подмножествам. Выбор какой-либо части данных производится по номерам указанных исследователем классов в соответствии со значениями характеристического вектора. При образовании новых классов (на этапе предварительной обработки или выделения их по графическим изображениям) заполняется вектор-индикатор (L_0 или L_{i+1} на рис. 1), по которому старое значение характеристического вектора корректируется в блоке формирования, а новое (W) поступает в блок выбора. Корректировка происходит и при выполнении указанных выше операций над классами. В первом варианте системы допускается одновременное образование до 36 классов (это число в принципе может быть увеличено).

Второй аспект организации данных — циркулирование их в системе. Так, Болл и Холл [3] указывают на необходимость автоматически пере-

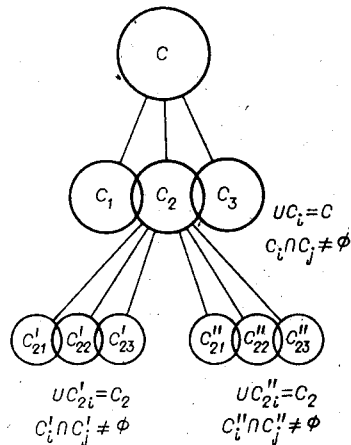


Рис. 2.

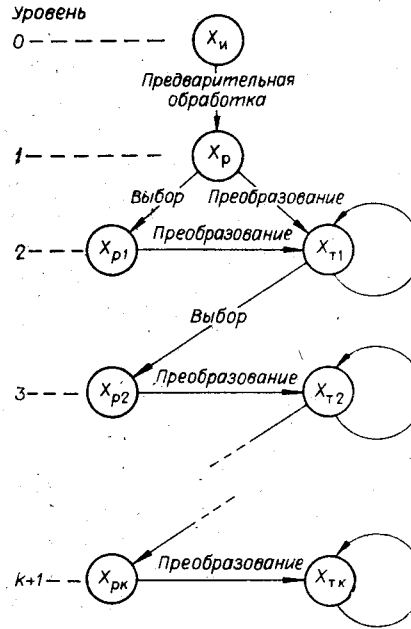


Рис. 3.

давать информацию из одного блока системы в другой, а Сэммон [4] использует для этой цели концепцию «текущих данных», когда реализация всех

основных возможностей системы не зависит от типа входных данных, будь то исходная совокупность, ее часть или преобразованное множество. Система «Вулканит» имеет многоуровневую организацию данных, схема которой показана на рис. 3. Здесь на нулевом уровне расположены исходные данные X_n , на первом — рабочие X_p , образующиеся из исходных на этапе предварительной обработки. Классифицированные данные X_{pi} являются частью преобразованных данных нижнего уровня $X_{t(i-1)}$, а соответствующие им X_{ti} формируются в результате очередного выполнения блока преобразования (возможно, несколько раз, что обозначено на рис. 3 циклами). Направления перемещений данных разных типов показаны на структурной схеме рис. 1. Попадая в управляющий блок, они становятся «текущими» до очередного выделения новых классов или перехода на более низкий уровень. Благодаря такой организации исследователь в любой момент имеет оперативный доступ ко всей полученной ранее информации. В первом варианте системы «Вулканит» $K=9$, чего обычно достаточно для решения практических задач, хотя ограничение и не является принципиальным.

Для эффективной работы с системой пользователь должен иметь представление о структуре данных; в самой же системе предусмотрен управляющий вектор, характеризующий ее текущее состояние. В нем фиксируется тип данных, с которыми в настоящий момент идет работа, номер выполняемого блока и информация о наличии классов. При изменении состояния — переходе к другому блоку или другому типу данных — этот вектор автоматически корректируется, а при осуществлении прерываний записывается конечным блоком на магнитную ленту. После очередного вызова системы он считывается и анализируется начальным блоком, который передает управление в то место, где было сделано прерывание.

Организация диалога. Известны две формы диалога. Для первой характерно то, что человек задает вопрос на специальном языке, а ЭВМ выбирает из набора имеющихся в ее распоряжении программ нужные для ответа и выполняет их в определенной последовательности [7]. Однако относительная сложность применяемого в таких случаях языка директив обычно существенно снижает эффективность диалога. Поэтому в большинстве систем анализа данных используется вторая форма [3, 4, 6], при которой сама машина после выполнения очередного

действия предлагает исследователю перечень возможных продолжений — «меню», а человек должен только указать интересующий его вариант. Таким способом достигается более тесный контакт с системой.

В системе «Вулканит» все вопросы разбиваются на две естественные группы: вопросы типа «разветвление» и «сообщение информации».

С помощью вопросов первого типа осуществляется собственно выбор нужного варианта продолжения работы из предлагаемого перечня. Так, для перехода из управляющего блока в блок отображения достаточно после появления вопроса:

1. СТАТИСТИЧЕСКИЙ АНАЛИЗ
2. ПРЕОБРАЗОВАНИЕ
3. ОТОБРАЖЕНИЕ
4. ПРОВЕРКА ГИПОТЕЗ

— задать третью возможность (указать цифру 3).

Вопросы типа «сообщение информации» отличаются как по назначению, так и по формам представления и ответа. Они могут возникать вслед за вопросами первого типа в том случае, когда для выбранного способа продолжения исследования необходима какая-нибудь дополнительная информация (набор параметров). Например, если в процессе предварительной обработки пользователь указал на удаление признаков, то на вопрос

НОМЕРА УДАЛЯЕМЫХ ПРИЗНАКОВ

можно ответить: 1, 5, 11—14, для того чтобы исключить из дальнейшего рассмотрения признаки с номерами 1, 5, 11, 12, 13, 14.

Внутри отдельных блоков вопросы организованы по иерархическому принципу таким образом, что, выбрав однажды более общую возможность, можно постепенно добраться до конкретных методов ее реализации. Фрагмент такой структуры условно изображен на рис. 4, где показана также одна особенность вопросов первого типа; при ответе на них допускается не только сообщение номера выбранного варианта, но и указание одного из трех системных кодов: возврат «В», прерывание «П», конец «К».

В первом случае происходит переход к вопросу типа «разветвление», находящемуся на более высоком уровне иерархии. Остальные ответы инициируют переход в конечный блок, где формируется соответствующий управляющий вектор.

Заметим, что присутствие оператора-предметника в системе диктует необходимость осуществления контроля правильности ответов исследователя. Однако этот вопрос более специальный и в данной работе не рассматривается.

Заключение. В настоящее время система «Вулканит» используется для обработки данных по силикатным составам вулканических пород Курило-Камчатской вулканической провинции [11].

Структура системы позволяет решать поставленную задачу последовательным применением различных возможностей за счет поступательного движения вперед и возвратов, изучения совокупности данных целиком и по частям, наглядного графического представления результатов.

Взаимосвязь исследователя с системой осуществляется с помощью

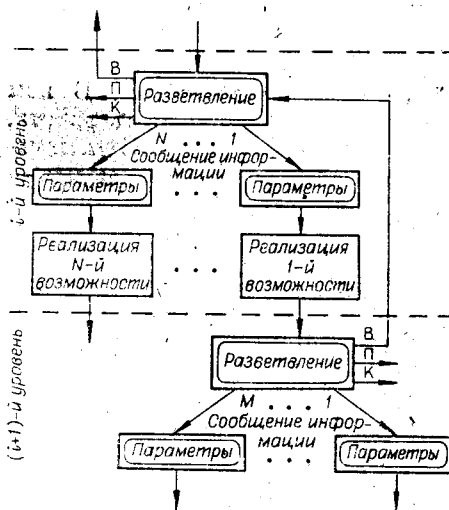


Рис. 4.

вопросов и ответов на естественном языке и проста в обращении. Пользователю для успешной работы достаточно знать лишь в общих чертах возможности выбираемых методов, а также иметь представление о типе данных, с которыми проводится исследование. Это открывает доступ к диалогу специалистам-предметникам, не знакомым с программированием и принципами работы ЭВМ.

ЛИТЕРАТУРА

1. Cooley W. W., Lohnes P. R. Multivariate Data Analysis.— N. Y.: John Wiley and Sons, 1971.
2. Канал Л. Обзор систем для анализа структуры образов и разработки алгоритмов классификации в режиме диалога.— В кн.: Распознавание образов при помощи цифровых вычислительных машин. М.: Мир, 1974.
3. Ball G. H., Hall D. J. Some Implications of Interactive Graphic Computer Systems for Data Analysis and Statistics.— Technometrics, 1970, vol. 12, N 1, p. 17—31.
4. Sammon J. W. Interactive Pattern Analysis and Classification.— IEEE Trans. on Computers, 1970, vol. C-19, N 7, p. 594—616.
5. Fisherkeller M. A., Friedman J. H., Tukey J. W. PRIM-9: Interactive Multidimensional Data Display and Analysis System.— Pacific 75.— Data: Use, Organiz. and Manag. Assoc. Comput. Mach. Reg. Conf. N. Y., 1975, p. 140—145.
6. Пакет прикладных программ ОТЭКС (версия 2М). (Обзор.) — Новосибирск, 1978.
7. Раудис Ш., Пикалис В., Юшкавичус К. Система оперативности разработки распознающих алгоритмов (СОРРА).— Статистические проблемы управления, 1978, № 27.
8. Горичкий Ю. А., Жаринов С. Е. Графическая диалоговая система «Вулканит» для классификации и анализа многомерных данных.— В кн.: VIII Всесоюз. сов. по проблемам управления. Таллин, 1980, кн. 2.
9. Hall D. J., Ball G. H., Wolf D. E. Applications of PROMENADE Data Analysis System.— IEEE Comput. Commun. Conf. Rec., 1969, p. 101—108.
10. Sammon J. W., Proctor A. H., Roberts D. F. An Interactive Graphic Subsystem for Pattern Analysis.— Pattern Recognition, 1971, vol. 3, p. 37—52.
11. Горичкий Ю. А., Жаринов С. Е., Масуренков Ю. П. Диалоговая система анализа данных и ее применение для классификации силикатных составов и вулканов Карымского центра.— В кн.: Вулканизм и вулканоструктуры (тезисы докладов). Тбилиси, 1980.

Поступила в редакцию 24 апреля 1981 г.

УДК 681.327.11

Е. Н. ВЛАДИМИРОВ, А. З. ПЛИСС, Л. З. ТАТКИН,
О. П. ЦЫБУЛЬСКИЙ
(Ленинград)

МНОГОКАНАЛЬНАЯ СИСТЕМА СБОРА ИНФОРМАЦИИ ПАРАЛЛЕЛЬНОГО ДЕЙСТВИЯ СО СКОРОСТНЫМ ПОСЛЕДОВАТЕЛЬНЫМ СЧИТЫВАНИЕМ

Известно, что детальное и более глубокое изучение сложных (широкополосных) процессов, сигналов и полей невозможно без использования систем анализа и измерительных систем с высоким разрешением и точностью [1, 2]. Однако чем выше разрешающая способность и точность системы и чем сложнее процесс (выше скорость его изменения), тем больше (при одноканальной системе) времени расходуется на обработку, так что система становится неприемлемой для анализа быстроизменяющихся процессов и принципиально непригодной для процессов с малым временем существования, например биологических. В таких случаях приемлем только многоканальный вариант системы сбора и обработки информации.

Теоретически минимальное количество дискретных каналов (датчиков) N_n , необходимых для воспроизведения любой функции, существую-