$$\sigma_l = \sqrt{\lim \sigma_l^2} = \sqrt{2/\pi}\,(2^2/n), \qquad (10)$$

поскольку $\lim\limits_{l\to\infty} \sum\limits_{i=0}^{l-1} \frac{1}{2^i} = 2$, а $\lim\limits_{l\to\infty} \sum\limits_{i=0}^{l-1} \frac{i+1}{2^i} = 2^4$.

Таким образом, в предельном случае, как это следует из формулы (10), измерения на $l$-м этапе будут обеспечивать скорость уменьшения среднеквадратической ошибки как $1/n$. Уменьшение числа этапов $l$, а также учет других обстоятельств, связанных с формированием эталонных траекторий, приводит к уменьшению достигаемого предела вплоть до значений $1/n^{3/4}$ для $l = 2$.

**Заключение.** Применение рекурсивных многоэтапных оценок измерения параметров повторяющегося сигнала позволяет увеличить точность измерения после проведения усреднения $n$ импульсов в $n^{3/4} - n$ раз, что дает дополнительный резерв повышения точности.

*Поступило в редакцию 29 декабря 1983 г.;*
*окончательный вариант — 7 июня 1984 г.*

---

УДК 519.95

**И. Ф. КЛИСТОРИН, Г. А. ТКАЧ, Г. Я. ШЕВЧЕНКО**

*(Кишинев)*

# ДИАЛОГОВАЯ СИСТЕМА ОБРАБОТКИ РЕЗУЛЬТАТОВ ЭКСПЕРИМЕНТОВ, ПРЕДСТАВЛЕННЫХ В ТАБЛИЧНОЙ ФОРМЕ

Многие исследовательские и прикладные задачи в области обработки разнотипных экспериментальных данных зачастую требуют обращения к целому ряду процедур как вычислительного, так и логического характера, выполняемых в определенной последовательности. Наблюдающееся усложнение этих задач (в частности, из-за увеличения размерности обрабатываемых данных) обусловливает обращение к ЭВМ для проведения машинного эксперимента и машинной обработки его результатов.

Появление сравнительно дешевых мини- и микроЭВМ со значительными возможностями создает предпосылки для автоматизации проведения машинного эксперимента и обработки данных с привлечением широкого круга пользователей. Однако в настоящее время для этих классов ЭВМ основным средством решения задач обработки данных является использование пакетов прикладных программ (ППП). В то же время простое применение ППП без его специальной организации — малоэффективное средство решения указанной проблемы.

В связи с этим актуальна разработка автоматизированных систем на базе мини- и микроЭВМ, позволяющих в режиме активного диалога проводить машинный эксперимент и обработку данных в условиях большой размерности массивов, разнотипности экспериментальных данных, необходимости запоминания, доступа к данным и быстрого манипулирования ими.

В работе рассматривается система, отвечающая перечисленным выше требованиям и реализующая один из возможных способов решения указанной проблемы. Решение заключается в табличном представлении данных, хранящихся в виде единого массива с фиксированной структурой, и в создании развитой модульной организации системы, позволяющей легко расширять ее возможности.

**Назначение.** Данная версия системы ДИСК (диалоговая система классификации) ориентирована на решение задач классификационной обработки разнотипных экспериментальных данных и обеспечивает: ввод данных с клавиатуры дисплея и запись их на диск; вызов в оперативную память файлов данных из архива, занесенного на диски; выполнение программ пользователя в автоматическом и диалоговом режимах; формирование архива результатов и запись его на диск; вывод на алфавитно-цифровой дисплей или АЦПУ исходных данных и результатов в табличной форме.

**Состав и структура.** Программное обеспечение системы ДИСК разработано на базе операционной системы РАФОС, предназначенной для эксплуатации СМ ЭВМ. В качестве языка программирования выбран язык Бейсик, обеспечивающий необходимый активный диалог с пользователем системы. Развитая модульная организация позволяет легко расширять возможности системы ДИСК и создавать необходимую ее конфигурацию. Ядром программного обеспечения служит управляющая программа ДИАЛОГ, на которую возложены функции ввода/вывода, доступа к данным, управления прохождением задач, диалог с пользователем, работа с библиотекой программ (БП) и т. п.

При работе системы в режиме диалога пользователь определяет вид одной из следующих процедур: ДАННЫЕ, РАБОТА, АРХИВ, ПЕЧАТЬ, ПОДСКАЗКА.

Процедура ДАННЫЕ формирует рабочий файл данных, включающий несколько блоков и имеющий фиксированную структуру. В каждом блоке записи следуют в строгом порядке: указываются параметры таблицы $N$, $M$, $K$, затем записывается сама таблица в виде двумерного массива

$$X(I, J), \quad I = \overline{1, N}, \; J = \overline{1, M+K}.$$

Здесь $N$ — число объектов в выборке; $M$ — число признаков; $K$ — число классов.

Рабочий файл данных TABL DAT формируется одним из следующих способов:

данные вводятся с клавиатуры дисплея в режиме диалога;

блок данных записывается в виде таблицы случайных чисел (в интервале $[a - b]$);

блок данных заносится в виде таблицы, состоящей из «0» и «1» при заданной вероятности появления «0» и «1»;

данные вводятся в оперативную память и поступают в рабочий файл с диска, где они хранились ранее.

Процедура РАБОТА организует выполнение программ пользователя. Система в режиме диалога запрашивает «план» работы: имена программ, их число, порядок выполнения, а также номер блока исходных данных в рабочем файле TABL DAT для каждой программы. Затем в оперативную память последовательно подгружается с диска каждая из указанных программ и выполняется в автоматическом режиме. Предусмотрена также возможность функционирования любой программы из БП в режиме диалога, что позволяет вести отладку новых алгоритмов.

Процедура ПЕЧАТЬ осуществляет распечатку архивных файлов, хранимых на диске. Система запрашивает информацию об имени файла и порядковом номере блока, содержащем интересующую пользователя таблицу. Распечатка производится в табличной форме. При необходимости распечатывается лишь нужный отдельный столбец или отдельная строка таблицы.

Процедура АРХИВ организует запоминание новых данных либо результатов выполнения программ в специально предназначенных для этого архивных файлах на диске. Пользователь должен указать имя, которое необходимо присвоить архивному файлу, а также рабочий файл, из которого следует брать данные для запоминания, т. е. либо TABL DAT в случае ввода и организации хранения новых данных, либо PROGR DAT при необходимости архивирования результатов обработки. Файл данных PROGR DAT формируется в ходе выполнения программ поблочно, причем последовательность блоков соответствует порядку прохождения задач.

Процедура ПОДСКАЗКА предусмотрена в качестве помощи пользователю и организует распечатку справочной информации о возможностях системы, а также о составе библиотеки программ.

На печать можно вывести по желанию каталог БП, пояснения к каждой программе, ссылки на литературу при использовании того или иного алгоритма, краткую инструкцию по эксплуатации системы.

**Обработка данных.** Обработка результатов экспериментов, представленных в табличной форме, проводится в системе ДИСК с помощью набора алгоритмов, образующих в совокупности библиотеку программ пользователя.

В зависимости от решаемых с помощью системы ДИСК задач пользователь выбирает из библиотеки требуемые алгоритмы, тем самым определяя конфигурацию системы.

БП размещается на дисках, и по мере необходимости та или иная программа вызывается по имени в оперативную память для выполнения.

В разработанной версии системы ДИСК БП построена на основе алгоритмов, реализующих логико-комбинаторный подход к обработке и классификации объектов. При этом состав БП обеспечивает выполнение следующих функций.

Программа ИНФОРМ вычисляет информативность как отдельных признаков $x_i$ $(i = \overline{1, M})$ (качественных или количественных) таблицы — обучающей выборки (ОВ), хранящейся в рабочем файле, так и произвольной совокупности признаков $x_{i_1}, \ldots, x_{i_\gamma}$ $(1 \leqslant \gamma \leqslant M)$ по предложенной ниже формуле:

$$W(x_{i_1}, \ldots, x_{i_\gamma}) = \frac{1}{K} \sum_{\Delta \in \Gamma} \max_m \left( \frac{b_\Delta^m}{b_m} \right), \qquad (1)$$

где $\Delta = t_{i_1}, t_{i_2}, \ldots, t_{i_\gamma}$ — произвольный набор значений признаков $x_{i_1}, \ldots, x_{i_\gamma}$; $b_\Delta^m$ — количество наборов ОВ из $m$-класса $(m = \overline{1, K})$, в которых $x_{i_j} = t_{i_j}$ $(j = \overline{1, \gamma})$; $t_{i_j}$ —

значения признака $x_{i_j}$ в наборе $\Delta$; $\Gamma$ — множество всех наборов значений признаков $x_{i_1}, \ldots, x_{i_\nu}$ в ОВ.

Оценка информативности признаков (1) основана на подходе, предложенном Кендаллом и Стьюартом [1] для получения оптимальных оценок связи категоризованных переменных. Отличительной особенностью (1) является унифицированный способ вычисления информативности качественных и количественных признаков. Для большего удобства и единообразия вычисления оценки информативности как отдельных признаков, так и произвольных их совокупностей в программе ИНФОРМ формула (1) реализована в следующем виде:

$$\widetilde{W}(x_i) = W(x_1,\ x_2,\ \ldots,\ x_{i-1},\ x_{i+1},\ \ldots,\ x_n). \qquad (2)$$

Это существенно облегчает проведение различных расчетов по оценке информативности признаков, экономит память и, кроме того, позволяет получить еще один вид оценки информативности признаков — по вкладу в информативность всего набора признаков (или какой-то его части).

Программа ПОИСК осуществляет по данным ОВ поиск информативных групп признаков (ИГП) в соответствии с заданным критерием. Поиск производится автоматически: по заданию пользователя отыскиваются либо все ИГП, либо только ИГП фиксированной длины.

Программа ДЕРЕВО организует построение по данным ОВ распознающего дерева (РД) в соответствии с алгоритмом, близким к описанному в [2]. Предварительно организуется выполнение программы ИНФОРМ, которая в этом случае округляет при необходимости вещественное значение признака до ближайшего целого значения, вычисляет информативность признаков ОВ и перед выполнением программы ДЕРЕВО ранжирует их по величине информативности. Результатом программы является минимизированное РД, которое, наряду с выявлением оптимального по составу набора признаков, является также решающим правилом при принятии решения о принадлежности распознаваемых объектов к тому или иному классу.

Принятие решения по дереву происходит по программе КЛАССИФИКАЦИЯ. Программа ЭКЗАМЕН предназначена для оценки точности (достоверности) РД. При выполнении этой программы данные из независимой экзаменационной выборки (ЭВ) проверяются на РД и вычисляется процент правильного их распознавания, что и служит основной точностью характеристики РД.

Программа ДООБУЧЕНИЕ организует перестройку РД в соответствии с данными ЭВ и результатами экзамена. Используется и ряд других программ, в том числе сервисных. В системе ДИСК предусмотрена возможность вывода исходных данных и результатов их обработки в табличной форме на алфавитно-цифровой дисплей или АЦПУ.

Выбранная табличная форма представления информации является весьма удобной и компактной при вводе, обработке, хранении данных и делает диалоговую систему достаточно универсальной для использования в задачах анализа, обработки и классификации разнотипных экспериментальных данных, представленных в табличной форме.

Для работы с системой пользователю не требуется специального ее знания, достаточно загрузить систему в ЭВМ, вызвать программу ДИАЛОГ и дать команду RUN — дальнейший ход работы определяется в режиме диалога.

Модульная организация системы позволяет легко расширять ее возможности посредством добавления новых процедур и/или изменения состава БП. Разработка подобного рода систем дает возможность эффективно проводить обработку и классификацию объектов различной природы.

## ЛИТЕРАТУРА

1. **Кендалл М., Стьюарт А.** Статистические выводы и связи.— М.: Наука, 1973.
2. **Василенко Ю. А., Роботишин В. И., Шевченко Г. Я.** Алгоритм обработки обучающих выборок большого объема.— В кн.: Проблемы бионики.— Харьков, 1979, вып. 23, с. 46—51.

*Поступило в редакцию 28 апреля 1984 г.*