

УДК 519.95:681.3.06

КЛАССИФИКАЦИЯ ЗАДАЧ ПРОГНОЗИРОВАНИЯ  
НА ТАБЛИЦАХ "ОБЪЕКТ-СВОЙСТВО"

Н.Г.Загоруйко

Экспериментальные данные, предназначенные для анализа с помощью математических методов и ЭВМ, обычно сводятся в протокол, который имеет вид прямоугольной матрицы  $M$ , строки которой соответствуют объектам  $\{a_i\}$ ,  $i = 1 \div m$ , а столбцы - признакам или свойствам  $\{x_j\}$ ,  $j = 1 \div n$ . На пересечении строки  $i$  и столбца  $j$  стоит символ  $b_{ij}$  (обычно - цифра), который обозначает результат взаимодействия прибора  $x_j$  с объектом  $a_i$ . Такую матрицу называют таблицей "объект-свойство" (ТОС). Если исходные данные представлены в другой форме (например, в виде граф-схем, графиков, текстов на естественном языке и т.д.), то обычно вначале решается задача сведения этих данных в ТОС - вычисляются структурные характеристики граф-схем, спектральные характеристики сигналов, 1-граммы текстов и т.д.

Все задачи, решаемые на ТОС, связаны 1) с поиском закономерных связей между разными частями таблицы и 2) с использованием найденных закономерностей для предсказания (прогнозирования) значений одних элементов таблицы по значениям других ее элементов. При этом, кроме фактов, содержащихся в ТОС, существенную роль играют некоторые дополнительные предположения, априорные знания, представленные в виде исходной эмпирической гипотезы  $h_0$  - в частности, предположения о виде закона распределения, о зависимости или независимости свойств, о типе шкал измерительных приборов, о классе закономерностей, в котором нужно искать интересующую нас закономерность и т.д. Сочетание всех возможных вариантов начальных условий  $h_0$ , конкретных данных в ТОС и целей анализа определяет разнообразие задач, решаемых на ТОС. Отметим среди них зада-

чи, связанные с прогнозированием. Классификацию этих задач будем представлять себе в виде иерархической системы (таблица).

Т а б л и ц а

Классификация задач прогнозирования

Семейство задач		1. Предсказание элементов $b_{i0}$ одного столбца			2. Предсказание элементов $b_{0j}$ одной строки			3. Предсказание элементов $b_{ij}$ разных столбцов и строк				
Классы задач		1.1. k=1	1.2. 1 < k < m	1.3. k=m	2.1. k=1	2.2. 1 < k < n	2.3. k=n	3.2. 1 < k < m · n	3.3. k=m · n			
Признаки		$\bar{x}$	$x_0$	$\bar{x}$	$x_0$	$\bar{x}$	$x_0$	$\bar{x}$	$x_0$	$\bar{x}$	$x_0$	$x_0$
Типы задач	И	*	*	*	*	*	*	*	*	*	*	*
	II	*	*	*	*	*	*	*	*	*	*	*
	K	*	*	*	*	*	*	*	*	*	*	*
	P	*		*		*		*		*	*	*
Количество типов задач		12		12		12		12		16		4

На верхнем уровне иерархии будем различать три семейства задач:

- 1) предсказание элементов  $b_{i0}$  одного и того же столбца ТОС;
- 2) предсказание элементов  $b_{0j}$  одной и той же строки ТОС;
- 3) предсказание элементов  $b_{ij}$ , находящихся в разных столбцах и строках ТОС.

Внутри семейства будем различать классы задач в зависимости от того, сколько элементов - один, несколько или все (k) нужно предсказать.

Типы задач выделяются в каждом классе в зависимости от того, в каких шкалах [ I ] измеряются предсказываемые элементы и в каких шкалах измерены остальные элементы, т.е. в зависимости от типа шкалы "целевого" признака  $x_0$  и "характеризующих" признаков  $\bar{x}$ .

На множестве всех возможных шкал будем отличать шкалы трех типов: номинальные (Н), порядковые (П), и так называемые количественные (К), куда будем относить шкалы разностей, интервалов, отношений и абсолютные. Если среди  $\bar{X}$  или  $x_0$  встречаются признаки, измеренные в разных шкалах, то их будем называть разнотипными и в таблице отмечалось знаком Р. Номер задачи отдельного типа состоит из номера семейства, класса и обозначения типа шкал для  $x_0$  и  $\bar{X}$ . Рассмотрим некоторые распространенные типы задач предсказания.

Если нужно прогнозировать один элемент  $b_{10}$  столбца  $x_0$ , измеренного в шкале наименований, то мы имеем дело с задачами "распознавания образов" следующих типов: I.I.H.H.; I.I.P.H.; I.I.K.H. и I.I.P.H.. Наиболее исследованной является задача I.I.K.H. - распознавания принадлежности объекта  $a_1$  к одному из заранее выделенных образов по значениям характеризующих признаков  $\bar{X}$ , измеренных в количественных шкалах [2]. Хорошо известна [2] также задача I.3.K.H. - задача таксономии, т.е. формирования признака  $x_0$ , который указывал бы для каждого объекта  $a_i$  имя его таксона. При этом все другие признаки  $\bar{X}$  этих объектов измерены в количественных шкалах, позволяющих использовать метрическую меру "близости", "расстояния" и т.п.

Промежуточное положение занимает задача I.2.K.H. - одновременного распознавания группы объектов по частично заданной обучающей выборке. Эта задача может решаться, например, методом таксономических решающих функций (алгоритм ТРФ [3]).

Если  $x_0$  измерен в шкале порядка, то мы должны решать задачи предсказания по свойствам объекта  $b_{10}$  его порядкового номера в ряду ранее упорядоченных объектов [4] - задачи I.2.H.П., I.2.П.П., I.2.K.П. и I.2.P.П.

В тех случаях, когда  $x_0$  измеряется в одной из более сильных (количественных) шкал, решаются задачи предсказания конкретного значения  $b_{10}$  или интервала числа, в котором может находиться это значение. Наиболее распространены в этом случае методы регрессионного анализа [5].

Аналогичный ряд задач имеет место и для элемента  $b_{0j}$  целевой строки  $a_0$ . Эта строка, например, может оценивать информативность признаков. Если это грубая оценка, указывающая лишь на принадлежность признака к одному из классов, то возникает задача распознавания, т.е. указания, к какому из этих классов относится некоторый новый признак. Если целевая строка оценивает признаки с

точностью до порядка, то мы будем иметь дело с задачей упорядочения, т.е. указания порядкового места данного нового признака среди ранее известных. При оценке информативности в более сильной шкале возникает задача количественной оценки элемента  $b_{0j}$ . Особый, слабо пока изученный класс задач составляют задачи в случае, когда элемент  $b_{0j}$  находится в строке, другие элементы которой заданы в разных шкалах (задачи 2.1.Н.Р.; 2.1.П.Р., 2.1.К.Р.)

Класс 2.2. типов задач будет иметь место для частично заданной целевой строки  $a_0$ , в которой требуется предсказать несколько элементов  $\{b_{0j}\}$ .

Новое семейство задач возникает в том случае, когда предсказываемые элементы  $\{b_{1j}\}$  расположены в разных строках и столбцах ТОС. Эти задачи могут решаться методами типа ЗЕТ и ВАНГА [3].

Можно указать на задачи класса 3.3. предсказания (генерирования) всех элементов ТОС, совокупность которых удовлетворяла бы некоторым заранее указанным в  $h_0$  свойствам. Примерами задач такого типа служат задачи генерирования таблиц случайных чисел с заданным законом распределения, таблиц данных в шкале порядка с заданным общим для таблицы групповым порядком, таблиц данных в номинальной шкале с заданным спектром имен и т.д. Эти таблицы бывают нужны для статистических испытаний различных алгоритмов обработки данных.

К настоящему времени разработаны методы для решения задач многих указанных типов. Ряд программ сведены в удобные для пользователей пакеты прикладных программ.

В пакете ОТЭКС [6], разработанном в Институте математики СО АН СССР совместно с Новосибирским государственным университетом, отделом отраслевого научно-технического прогнозирования ИПСС СССР, с участием некоторых других организаций, включены программы, решающие около половины типов задач таблицы. Пакет содержит около 20 программ. Библиотеку этих функциональных модулей дополняет системная часть, которая позволяет непрограммирующему пользователю формулировать задачу на языке понятий из своей прикладной области.

Из нерешенных задач обработки данных на ТОС наиболее актуальными являются следующие:

1. Разработка методов обработки ТОС с разнотипными признаками.

2. Разработка методов обработки ТОС больших размеров (по сравнению с оперативной памятью).

3. Разработка полигонов для испытания и сравнения алгоритмов решения задач всех типов.

#### Л и т е р а т у р а

1. СЛШЕС П., ЗИНЕС Дж. Основы теории измерений. -В кн.: Психологические измерения. -М., 1967, с.9-110.
2. ЗАГОРУЙКО Н.Г. Методы распознавания и их применение. -М.: Сов.радио, 1972. - 306 с.
3. ЗАГОРУЙКО Н.Г. Эмпирическое предсказание. - Новосибирск: Наука, 1980. - 123 с.
4. ВОРОНИН Ю.А., ГОРЕЛОВА Н.Г. Об автоматизированной системе для постановки и решения задачи упорядочения. -В кн.: Системы и методы обработки данных, Новосибирск, 1978, с. 3-21.
5. ДРЕЙШЕР Н., СМИТ. Прикладной регрессионный анализ. - М.: Статистика, 1973. - 447 с.
6. ЗАГОРУЙКО Н.Г., ЛБОВ Г.С., МАШАРОВ Ю.П. Пакет прикладных программ для обработки таблиц экспериментальных данных ОТЭК-1. -В кн.: Вопросы обработки информации при проектировании систем (Вычислительные системы, вып. 69.) Новосибирск, 1977, с. 93-102.

Поступила в ред.-изд.отд.  
10 апреля 1981 года