ДИСКРЕТНЫЙ АНАЛИЗ

Сборник трудов

1968 г. Института математики СО АН СССР Выпуск I2

УДК 519.95

НЕКОТОРЫЕ ТАБЛИЧНЫЕ ЧИСЛА

А.Н. Дмитриев

Пусть задан ряд объектов $M=\{\tilde{\alpha}_1,\tilde{\alpha}_2,\dots,\tilde{\alpha}_K\}$, которые охарактеризованы набором из n предикатов $X_1,\dots,X_n$, причем для некоторых объектов некоторые предикаты неизвестны.

Заданы также "целевые предикаты" $X_{n+1},\dots,X_{n+r}$, которые сохраняют постоянное значение для объектов $\tilde{\alpha}_1\tilde{\alpha}_2,\dots,\tilde{\alpha}_K$ и определяют "тип" – т.е. множество объектов, для которых

$$X_{n+1}\equiv\gamma_1,\dots,X_{n+r}=\gamma_{n+r}.$$

В работе [I] были введены такие понятия, как информационный вес признака, информационный вес строки. Там же было указано, каким образом знание этих чисел позволяет классифицировать объекты из M и прогнозировать принадлежность элементов из СМ[ж)] к тому же типу (по целевым предикатам $X_{n+1},\dots,X_{n+r}$), что и элементы M.

ж) СМ – совокупность объектов, подлежащих диагностике.

§ I. Априорные табличные числа

Ниже мы введем ряд табличных чисел, которые позволяют выявить интересные свойства объектов из M.

Пусть задана таблица T, которую в дальнейшем мы будем называть основной.

	X_1	X_2	$\cdots$	X_n
$\tilde{\alpha}_1$	α_{11}	α_{12}	$\cdots$	α_{1n}
$\tilde{\alpha}_2$	α_{21}	α_{22}	$\cdots$	α_{2n}
$\vdots$				
$\tilde{\alpha}_K$	α_{K1}	α_{K2}	$\cdots$	α_{Kn}

Здесь α_{ij}, $1\le i\le n$, $1\le j\le K$, принимают одно из трех значений 0, I, –.

Таблицы, в которые входят лишь объекты со всеми известными целевыми предикатами назовем эталонными, с неизвестными – таблицами проб. Мы будем рассматривать только допустимые [I] таблицы.

ОПРЕДЕЛЕНИЕ I. Число $\xi=\frac{\psi_0}{\psi_1}$, где ψ_0, ψ_1 есть число символов α_{ij}, равных соответственно нулю или единице, называется пропорциональностью таблицы. При изучении таблиц, соответствующих реальным объектам, пропорциональность таблиц обычно близка к единице.

ОПРЕДЕЛЕНИЕ 2. Число $\nu=\frac{\psi_0+\psi_1}{K\cdot n}$ называется изученностью таблицы T. Если все объекты таблицы принадлежат по целевым признакам к одному типу L, то ν называется также изученностью типа L.

Пусть в строке, соответствующей объекту $\tilde{\alpha}_i$, число нулей равно X_i^0 и число единиц – X_i^1.

Тогда число $\nu^i=\frac{X_i^0+X_i^1}{n}$ назовем изученностью объекта $\tilde{\alpha}_i$ в таблице T.

Таблицы, для которых $\nu=1$, составлены из полностью изученных по признакам $X_1,\dots,X_n$ таблиц.

Обычно для эталонных таблиц $\nu=1$, а для таблиц проб:

$$0<\nu<1.$$

Пусть $X_{i_1},\dots,X_{i_m}$ – все предикаты из набора $X_1,\dots,X_n$, которые сохраняют на всех элементах $\tilde{\alpha}_1,\dots,\tilde{\alpha}_K$ постоянное значение, равное нулю или единице. Признаки $X_{i_1},\dots,X_{i_m}$ назовем отождествляющими признаками таблицы T (типа L); осчто и элементы M.

тальные признаки из числа $X_1, \ldots, X_n$ назовем различающими.

Будем говорить, что предикаты $X_{i_1}, \ldots X_{i_m}, X_{n+1}, \ldots, X_{n+r}$ образуют критерий общности таблицы T (типа L).

ОПРЕДЕЛЕНИЕ 3. Число $\ell = \frac{m+r}{n-m}$ назовем родством объектов таблицы T (типа ℓ).

Заметим, что $\ell > 0$. Число $n-m$ не обращается в нуль, так как в этом случае таблица T не является допустимой.

Числа, введенные определениями I-3, назовем априорными числами. Они легко вычисляются по таблице T.

§ 2. Апостериорные табличные числа

Пусть на допустимой таблице вычислены информационные веса $P(i)$ различающих признаков и информационные веса $\mathcal{J}(S)$ объектов (строк) S [I]. Пусть также признаки и строки упорядочены по убыванию величин $P(i)$ $\mathcal{J}(S)$ соответственно, т.е. в исходной таблице соответствующим образом переставлены строки и столбцы (построена апостериорная таблица).

Рассмотрим некоторый условный объект S, которому соответствует строка со значениями всех различающих признаков, равными I.

Пусть в таблице вес $P(i)$ вычислен для признаков $X_1, \ldots, X_m$.

Имеем: $\mathcal{J}(S) = \sum_{i=1}^{m} P(i)$.

Пусть S_t – произвольная строка таблицы.

ОПРЕДЕЛЕНИЕ 4. Величину

$$\nu(S_t) = \frac{\mathcal{J}(S) - \mathcal{J}(S_t)}{\mathcal{J}(S)} = 1 - \frac{\mathcal{J}(S_t)}{\mathcal{J}(S)}$$

назовем коэффициентом типичности объекта.

Очевидно, $0 \leq \nu(S_t) \leq 1$.

Для ряда задач [3] был обнаружен следующий факт: чем меньше коэффициент $\nu(S_t)$, тем более ценным (по данным целевым признакам) является объект S_t.

Значения коэффициентов типичности для всей совокупности объектов, заданных таблицей T (типом L) имеют обычно ярко выраженные разрывы, т.е. тип L по коэффициентам типичности разбивается на подтипы. Для объектов одного типа колебание коэффициента типичности невелико, для объектов из разных подтипов оно существенно больше.

Разбиение объектов на подтипы по величине $\nu(S_t)$ позволяет ввести характеристику, описывающую степень разложимости типа на подтипы.

Зададим каждый подтип отдельной таблицей и измерим "расстояние" между двумя подтипами L_1 и L_2.

Пусть подтипу L_1 принадлежат строки $S_1', S_2', \ldots, S_K'$ подтипу L_2 – строки $S_1^2, \ldots, S_t^2$.

Положим

$$L_1[\mathcal{J}(S)] = \frac{1}{K}\left(\sum_{i=1}^{K} \mathcal{J}(S_i')\right) \cdot \frac{1}{\sum_{i=1}^{m} P(i)},$$

$$L_2[\mathcal{J}(S)] = \frac{1}{t}\left(\sum_{i=1}^{t} \mathcal{J}(S_i^2)\right) \cdot \frac{1}{\sum_{i=1}^{m} P(i)}.$$

ОПРЕДЕЛЕНИЕ 5. Величину

$$\rho = 1 - \frac{L_1[\mathcal{J}(S)]}{L_2[\mathcal{J}(S)]}$$

назовем коэффициентом разложимости на подтипы L_1 и L_2.

Введем также величину, оценивающую степень близости объектов в типе.

Пусть $\tilde{\mathcal{J}}(S_i) = \frac{\mathcal{J}(S_i)}{\sum_{i=1}^{m} P(i)}$, K – число строк таблицы,

$$\max_{1 \leq i \leq K} \tilde{\mathcal{J}}(S_i) = (\tilde{\mathcal{J}}(S_i))_{max}, \quad \min_{1 \leq i \leq K} \tilde{\mathcal{J}}(S_i) = (\tilde{\mathcal{J}}(S_i))_{min}.$$

ОПРЕДЕЛЕНИЕ 6. Число $\varkappa = \frac{\sum_{i=1}^{K} \tilde{\mathcal{J}}(S_i)}{K[(\tilde{\mathcal{J}}(S))_{max} - (\tilde{\mathcal{J}}(S))_{min}]}$ назовем мерой родства объектов таблицы (типа).

Величину $\varkappa$ полезно вычислять и для каждого из выделенных подтипов. Мера родства объектов в типе обычно указывает на качество формирования типа (при априорном задании объектов). Если нет предварительного указания о разбиении объектов на подтипы, выделение "лучшего" подтипа оказывается полезным при формировании "идеального объекта" исследуемого типа.

Величины ρ и $\varkappa$ оказываются полезными при решении некоторых диагностических задач [4].

Литература

1. А.Н. Дмитриев, Ю.И. Журавлев, Ф.П. Кренделев. О математических принципах классификации предметов и явлений. Сб. "Дискретный анализ", № 7, 1967.
2. А.Н. Дмитриев, Ю.И. Журавлев, Ф.П. Кренделев. Об одном принципе классификации и прогноза геологических объектов и явлений. Геология и геофизика, № 5, 1968.
3. Ф.П. Кренделев, А.Н. Дмитриев, Ю.И. Журавлев. Сравнение геологического строения зарубежных месторождений докембрийских конгломератов с помощью дискретной математики. ДАН СССР, № 5, том 173.
4. А.Н. Дмитриев, В.В. Золотухин, Ю.Р. Васильев. Опыт применения дискретной математической обработки информации по дифференцированным трапповым интрузиям Северо-Запада Сибирской платформы. Советская геология (в печати 1967).

Поступила в редакцию
20.II.1968 г.

ДИСКРЕТНЫЙ АНАЛИЗ

Сборник трудов

1968 г. Институт математики СО АН СССР Выпуск 12

УДК 519.95

О ТУПИКОВЫХ И МИНИМАЛЬНЫХ ТЕСТАХ ДЛЯ ОДНОГО КЛАССА ТАБЛИЦ

В.Н. Носков

§ 1. Введение

Пусть $\{T\}_{nm}$ – множество таблиц, состоящих из m попарно различных строк и n столбцов, заполненных символами 0 и 1.

Набор столбцов $\langle j_1, j_2, \dots, j_k \rangle$ таблицы $T \in \{T\}_{nm}$ называется тестом[x)] для T, если после удаления из T всех столбцов, не вошедших в число $\langle j_1, j_2, \dots, j_k \rangle$, все строки оставшейся таблицы будут различны.

Тест называется тупиковым, если после удаления из него любого столбца он перестает быть тестом.

Число столбцов в тесте называется длиной теста. Тест наименьшей длины среди всех тестов таблицы называется минимальным.

x) Более общее определение теста дано в [1].