Serdica

Bulgariacae mathematicae publicationes

Сердика

Българско математическо списание

ПРОСТРАНСТВО НЕУПОРЯДОЧЕННЫХ КАЧЕСТВЕННЫХ ПРИЗНАКОВ НОМИНАЛЬНОГО И АДДИТИВНОГО ТИПА

ГЕОРГИ СТ. НИКОЛОВ

Создание математических моделей для обработки качественных данных связано с трудностями, возникающими в связи с возможностью интерпретирования и смысленности использования математических операций, со способами математической записи собранной информации и с непокиданием совокупности определенного вида нечисловых признаков при построении одной или другой модели (аппроксимационной). Возможно, самая простая совокупность нечисловых признаков, в которой основные математические операции дают признак из этой совокупности и аппроксимационные модели имеют хорошие свойства, это совокупность номинальных и аддитивных признаков. В настоящей работе определены основные математические операции в этой совокупности и доказаны некоторые свойства ее элементов.

1. Определение и свойства качественных признаков без упорядочения. Пусть дано множество Ω, состоящее из N объектов $\omega_1, \omega_2, \ldots, \omega_N$.

Определение 1. *Будем говорить, что R — качественный признак без упорядочения, определенный на множестве Ω, если разбивает множество Ω на c подмножеств (категорий) $A_1, A_2, \ldots, A_c$, причем $1 \leqq c \leqq p$ (p — конечное) $, \bigcup_{i=1}^{c} A_i \Omega$, и не существует даже и частичное упорядочение между категориями.*

Когда $A_i \bigcap_{i \neq j} A_j = \Phi$ для каждого $i, j = 1, 2, \ldots, c$, признак называется номинальным. При нарушении этого условия хотя бы для одной пары категорий признак называется аддитивным.

Матрица связей r качественного признака без упорядоченности R — это булевая матрица размерности $N \times N$, причем $r_{ij} = 0$, когда объекты ω_i и ω_j не принадлежат одновременно ни одной из категорий $A_1, A_2, \ldots, A_c$, $r_{ij} = 1$, когда оба объекта принадлежат хотя бы одной категории вместе (без значения, которой и сколько) и $r_{ii} = 1$ для $i = 1, 2, \ldots, N$, поскольку каждый элемент принадлежит вместе с собой одной категории.

При нормальных признаках посредством подходящего одновременного размещения рядов и столбцов матрица r может быть приведена в блочно-диагональный вид [2], а при аддитивных признаках этого невозможно сделать. Номинальные и аддитивные признаки имеют различные свойства, поэтому через X будем обозначать номинальные, а через L — аддитивные признаки, а когда рассматриваемый признак может быть признаком двух видов, будем обозначать его через R.

Обозначим через $\Omega \times \Omega$ множество всех упорядоченных пар (ω_i, ω_j) элементов Ω. Каждое подмножество ρ множества $\Omega \times \Omega$ называется бинарным отношением [1].

Через $R(\omega_i)$ обозначим категорию, которой принадлежит объект ω_i. Существует следующая связь между бинарным отношением и признаком R [2]:

Теорема 1. *Для любого ρ существует R и для любого R существует ρ, так что $\rho = \{(\omega_i, \omega_j) \mid R(\omega_i) = R(\omega_j)\}$.*

Отношение ρ для признака R — рефлексивно: т. е. $(\omega_i, \omega_j) \in \rho$ для любого $\omega_i \in \Omega$, и симметрично, т. е. если $(\omega_i, \omega_j) \in \rho$, то $(\omega_j, \omega_i) \in \rho$.

При номинальных признаках ρ является и транзитивным, т. е. если

$$[(\omega_i, \omega_j) \in \rho] \wedge [(\omega_j, \omega_k) \in \rho] \Rightarrow (\omega_i, \omega_k) \in \rho. \tag{1}$$

Для аддитивных признаков транзитивное свойство не выполняется, поскольку $R(\omega_i)$ может принимать несколько значений, но для всех элементов одной категории транзитивность выполняется.

Между элементами матрицы r „объект$\times$объект" и бинарным отношением ρ существует узкая связь, т. е. $r_{ij}=1$, когда $(\omega_i, \omega_j) \in \rho$ и $r_{ij}=0$ при $(\omega_i, \omega_j) \notin \rho$. Используя транзитивность и симметричность отношения ρ, можно сказать, что матрица связей r признака R-симметрична с единицами по главной диагонали, а условие (1) транзитивности ρ выражается с помощью элементов матрицы r следующим образом: бинарное отношение ρ транзитивно, если из

$$r_{ij} \,.\, r_{ik} = 1 \Rightarrow r_{jk} = 1. \tag{2}$$

Каждая симметрическая булевая матрица с единицами по главной диагонали размерности $N \times N$, для элементов которой выполняется (2), определяет однозначно один номинальный признак X, и наоборот, каждый номинальный признак определяет одну матрицу с верхними свойствами.

Для аддитивных признаков это несправедливо. При построении матрицы r теряется информация, т. е. если ω_i и ω_j принадлежат вместе одной или нескольким категориям, то всегда будем писать $r_{ij}=1$. Если этого не пишем, то это приводит к утяжеленной записи r (она не будет булевой матрицей), и трудно будет проводить действия с аддитивными признаками. Поэтому введем понятие „базовое" множество и с его помощью построим однозначно по данной матрице связей r систему базовых множеств, определяющую аддитивный признак.

Определение 2. *Будем называть B базовым множеством признака R, если оно содержит все элементы Ω, для которых условие* (2) *выполняется в совокупности.*

Через $B_1, B_2, \ldots, B_c$ будем обозначать базовые множества, которые определяют признак R. Используя определение базового множества, можно сказать, что включение $B_i \subset B_j$ невозможно ни для какого $i, j = 1, 2, \ldots, c$, поскольку если это включение выполняется для некоторых i и j, то B_i не содержит максимального числа элементов Ω, для которых выполняется (2), и, следовательно, не является базовым множеством.

Замечание. Категории одного номинального признака являются базовыми множествами для него.

По заданной матрице связей r можно однозначно определить систему базовых множеств для аддитивного признака L. Если он задан посредством категорий, на которые разбивает Ω, т. е. знаем $B_1, B_2, \ldots, B_c$ при условии, что исключим те, для которых выполнено $B_i \subset B_j$, получим систему базовых множеств для L.

Теорема 2. *Любые две различные $N \times N$ симметрические матрицы с единицами по главной диагонали определяют признаки, которые имеют хотя бы по одному несовпадающему базовому множеству.*

Доказательство. Обозначим через r_1 и r_2 матрицы, а через $B_1^1, B_2^1, \ldots, B_{c_1}^1$ и $B_1^2, B_2^2, \ldots, B_{c_2}^2$ — базовые множества, которые они определяют.

Поскольку по условию обе матрицы различные, то хотя бы для одного элемента будет выполняться $r_{ij}^1 \neq r_{ij}^2$, где $\{r_{ij}^1\}_{i,j=1}^N$ и $\{r_{ij}^2\}_{i,j=1}^N$ — элементы соответственно матриц r_1 и r_2. Это означает, что одно из чисел — ноль, а другое — единица. Пусть $r_{ij}^1=0$ и $r_{ij}^2=1$, следовательно, ω_i и ω_j не принадлежат ни одному из базовых множеств $B_1^1, B_2^1, \ldots, B_{c_1}^1$ и принадлежат хотя бы одному, обозначим его через B_p^2 из системы $B_1^2, B_2^2, \ldots, B_{c_2}^2$. Отсюда получаем, что $B_p^2 \neq B_i^1$ для $i=1, 2, \ldots, c_1$, чем теорема доказана.

2. Свойства базовых множеств. Теорема 3. *Число базовых множеств для аддитивного признака L, определенного на множестве Ω, имеющего N объектов* ($N>4$), *не может быть больше, чем*: 1) $N^2/4$ *при четном* N; 2) $(N^2-1)/4$ *при нечетном* N.

Доказательство. Конструируем признак L с матрицей связей r, которая: 1) содержит максимальное число единиц и для которой 2) ни для каких трех элементов не выполняется условие (2).

Из второго условия следует, что каждая единица над главной диагональю матрицы r определяет двухэлементное базовое множество, а из того, что выполнено и условие 1) — что L — признак, имеющий максимальное число базовых множеств.

Для того, чтобы нарушалось условие (2), достаточно, чтобы из соотношения $r_{\mu j} \cdot r_{\mu k}=1$ следовало, что $r_{jk}=0$ ($\mu, j, k=1, 2, \ldots, N$). Если возьмем

$$r_{\mu j}=\begin{cases}1 & \text{при } \mu, j \text{ — одной и той же четности}\\ 0 & \text{при } \mu, j \text{ — разной четности,}\end{cases} \tag{3}$$

то ни для каких трех элементов матрицы r условие (2) не будет выполнено.

Имея в виду (3), матрицу r можно записать

$$r=\left\|\begin{matrix}1 & 1 & 0 & 1 & 0 & 1 & \ldots & 1 & 0 & 1\\ & 1 & 1 & 0 & 1 & 0 & \ldots & 0 & 1 & 0\\ & & 1 & 1 & 0 & 1 & \ldots & 1 & 0 & 1\\ & & & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot\\ & & & & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot\\ & & & & & & & 1 & 1 & 0\\ & & & & & & & & 1 & 1\\ & & & & & & & & & 1\end{matrix}\right\|$$

Добавим еще одну единицу в r, например, $r_{ij}=1$, где i и j — одной и той же четности. Тогда условие транзитивности (2) выполнено для каждого μ четности, отличной от четности i, т. е. $r_{\mu i} \cdot r_{\mu j}=1 \Rightarrow r_{ij}=1$ и элементы ω_μ, ω_i и ω_j будут принадлежать одному и тому же базовому множеству. Меняя μ (оно может принимать $N/2$ различных значений), получим $N/2$ базовых множеств с тремя элементами, причем для образования каждого из них используются две единицы из r и новая — добавленная. Следовательно, добавлением только одной единицы в r число базовых множеств уменьшается на $N/2$.

I. Пусть N — четное. Число единиц S над главной диагональю матрицы r равняется $S=1+3+5+\cdots+(N-1)=N^2/4$.

II. Пусть N — нечетное. $S=2+4+6+\cdots+(N-1)=(N^2-1)/4$.

Замечание. Единицы в матрице r могут быть расположены и другим образом, так чтобы условие (2) нарушалось, но в ней не могут быть поставлены больше, чем $N^2/4$ единиц.

Теорема 4. *Если матрица связей аддитивного признака L содержит над главной диагональю $N^2/4$ (N — четное) или $(N^2-1)/4$ (N — нечетное) единиц ($N>2$), то в зависимости от их расположения в матрице число базовых множеств для признака L меняется на отрезках:*

1. $[2, N^2/4]$ — *при четном* N,
2. $[2, (N^2-1)/4]$ — *при нечетном* N.

Доказательство 1. Для того чтобы все объекты Ω попали в одно базовое множество, необходимо $(N^2-N)/2$ единиц над главной диагональю, и поскольку $N^2/4<(N^2-N)/2$, то признак L не может иметь только одно базовое множество.

Образуем базовые множества $B_1=\{\omega_1, \omega_2, \ldots, \omega_{N/2+1}\}$ и $B_2=\{\omega_{N/2+1}, \ldots, \omega_N\}$. Любые два элемента, принадлежащие одному и тому же множеству, определяют одну единицу в матрице связей r. Тогда общее число единиц равняется

$$S=\binom{N/2+1}{2}+\binom{N/2}{2}=\frac{(N/2+1)!}{2!\,(N/2+1-2)!}+\frac{(N/2)!}{2!\,(N/2-2)!}=\frac{1}{2}[N/2(N/2+1)+N/2(N/2-1)]$$
$$=N^2/4.$$

Следовательно, система базовых множеств может содержать не меньше, чем два множества, а в теореме 3 доказали, что их максимальное число — $N^2/4$.

Аналогично доказывается и утверждение для второго сегмента, беря $B_1=\{\omega_1, \omega_2, \ldots, \omega_{(N+1)/2}\}$ и $B_2=\omega\{_{(N+1)/2}, \omega_{(N+1)/2}+1, \ldots, \omega_N\}$.

3. Действия с признаками. Определение 3. *Сечением двух признаков $R_1\cap R_2$, заданных посредством своих базовых множеств, будем называть признак R_3, базовые множества которого получаются из всевозможных сечений базовых множеств R_1 и R_2.*

Если признаки R_1 и R_2 имеют соответственно матрицы связей r_1 и r_2, то определение сечения можно дать следующим образом: это признак R_3 с матрицей связей r_3, где $r_3=\| r_{ij}^2\|_{i,j=1}^N$; $r_{ij}^3=\min(r_{ij}^1, r_{ij}^2)$.

Сечением двух матриц r_1 и r_2 будем называть матрицу r_3, образованную вышеуказанным способом.

Обозначим через $\pi(R)$ множество всех качественных признаков без упорядоченности. Используя второе определение сечения двух признаков, можем сказать, что оно дает также признак из $\pi(R)$, так как сечение двух симметрических матриц с единицами по диагонали — это матрица с теми же свойствами. Аналогично, сечение двух номинальных признаков дает номинальный [2], а для остальных двух случаев (два аддитивных; аддитивный и номинальный) полученный признак может быть какого-нибудь одного из двух видов.

Определение 4. *Объединением $R_1\cup R_2$ признаков будем называть признак R_3, два объекта которого принадлежат хотя бы одному его базовому множеству, если они находятся хотя ды в одном базовом множестве первого или второго признака.*

На языке матриц связей это признак R_3, заданный матрицей

$$r_3=\| r_{ij}^3\|_{i,j=1}^N;\quad r_{ij}^3=\max(r_{ij}^1, r_{ij}^2).$$

Полученную матрицу r_3 будем обозначать $r_3=r_1\cup r_2$.

Определение 5. *Разностью $R_1 - R_2$ признаков будем называть признак R_3, два объекта которого принадлежат одному базовому множеству, если они вместе находятся хотя бы в одном базовом множестве R_1 и ни в одном R_2.*

Используя матрицы связей, получаем, что R_3 — это признак, определенный матрицей

$$r_3 = \| r_{ij}^3 \|_{i,j=1}^N ; \quad r_{ij}^3 = \begin{cases} \min(r'_{ij}, \, | r'_{ij} - r_{ij}^2 |) & \text{при } i \neq j \\ 1 \quad \text{при} \quad i = j. \end{cases}$$

Будем записывать, что $r_3 = r_1 - r_2$.

Определение 6. *Симметрической разностью $R_1 \triangle R_2$ будем называть признак R_3, определенный следующим образом:*

$$R_3 = R_1 \triangle R_2 = (R_1 \cup R_2) - (R_1 \cap R_2) = (R_1 - R_2) \cup (R_2 - R_1).$$

Определение 7. *Дополнением $\overline{R}$ признака R будем называть признак, два объекта которого принадлежат хотя бы одному базовому множеству $\overline{R}$, когда они не принадлежат ни одному базовому множеству признака R.*

$\overline{R}$ можем определить с помощью матрицы $\overline{r} = I - r + E$, где матрица I составлена только из единиц.

Определение 8. *Включение признака R_1 в R_2 ($R_1 \subseteq R_2$) имеем, когда каждая пара элементов, принадлежащая хотя бы одному базовому множеству ирзнака R_1, принадлежит хотя бы одному базовому множеству признака R_2.*

Используя матрицы связей, определение принимает вид $R_1 \subseteq R_2$, если $r_1 \cap r_2 = r_1$.

Для дальнейших вычислений удобно ввести операцию включения и для матриц, т. е. $r_1 \subseteq r_2$, когда $r_1 \cap r_2 = r_1$.

4. Метрика в $\pi(R)$. Будем использовать расстояние, определенное Миркиным [2]:

$$d(R_1, R_2) = \sum_{i,j=2}^{N} | r_{ij}^1 - r_{ij}^2 |.$$

Используя определение разности двух признаков, находим

$$\sum_{i,j=1}^{N} | r_{ij}^1 - r_{ij}^2 | = |(r_1 \cup r_2) - (r_1 \cap r_2)| - N = | r_1 \triangle r_2 | - N,$$

где через $|r|$ обозначено число единиц матрицы r. Следовательно,

$$d(R_1, R_2) = | r_1 \triangle r_2 | - N. \tag{4}$$

Для номинальных признаков это не совсем подходяще, поскольку если имеем два признака X_1 и X_2, которые отличаются только по тому, что один элемент из A_S-ой категории переброшен в A_P-ую, то расстояние между ними дается формулой $d(X_1, X_2) = 2[N_S - 1 + N_P]$. Отсюда можно сделать вывод, что если перебрасывание сделано между категориями с большим числом элементов, то расстояние между почти идентичными признаками в некоторых случаях будет очень велико.

При аддитивных признаках мы заметили, что добавление одной единицы в матрице связей ведет или к увеличению базовых множеств, или к их уменьшению (при теореме 3 число базовых множеств уменьшается на $N/2$, причем в $N/2$ базовых множеств число элементов увеличивается на один), т. е. получаем, что для признаков с расстоянием между ними, равным двум, будут большие различия между числом их базовых множеств, а также и между числом элементов в них.

Теорема 5. *Если для признаков R_1, R_2 и R_3 выполняется включение $R_1 \subset R_2 \subset R_3$, то $d(R_1, R_3) = d(R_1, R_2) + d(R_2, R_3)$.*

Доказательство. Из третьей аксиомы о расстоянии имеем

$$d(R_1, R_3) \leq d(R_1, R_2) + d(R_2, R_3), \tag{5}$$

а из (4) и $r_1 \subset r_2 \subset r_3$ записываем

$$d(R_1, R_3) = |r_1 \triangle r_3| - N = |(r_1 \cup r_3) - (r_1 \cap r_3)| - N = |r_3 - r_1| - N;$$

$$d(R_1, R_2) = |r_1 \triangle r_2| - N = |r_2 - r_1| - N;$$

$$d(R_2, R_3) = |r_2 \triangle r_3| - N = |r_3 - r_2| - N.$$

Подставляя в (5), находим

$$|r_3 - r_1| \leq |r_2 - r_1| + |r_3 - r_2| - N. \tag{6}$$

Чтобы доказать теорему, достаточно доказать неравенство

$$|r_2 - r_1| + |r_3 - r_2| - N \leq |r_3 - r_1|, \tag{7}$$

для чего покажем, что каждому месту ненулевого элемента в матрицах $(r_2 - r_1)$ и $(r_3 - r_2)$ соответствует ненулевой элемент на том же месте в матрице $(r_3 - r_1)$ и что единичные элементы матриц $(r_2 - r_1)$ и $(r_3 - r_2)$ находятся на различных местах за исключением элементов по диагонали.

Пусть элемент r_{ij} матрицы $(r_2 - r_1)$ — единица $(i \neq j)$. Это означает, что $r^2_{ij} = 1$ и $r^1_{ij} = 0$, но $r_2 \subset r_3$, следовательно, $r^3_{ij} = 1$. Из $r^1_{ij} = 0$ и $r^3_{ij} = 1$ находим, что i, j-ый элемент матрицы $(r_3 - r_1)$ также отличен от нуля, а из того, что $r^1_{ij} = 1$, получаем, что i, j-ый элемент матрицы $(r_3 - r_2)$ — ноль. Аналогичным способом доказывается, что ненулевые элементы и матрицы $(r_3 - r_2)$ совпадают по местам с теми же матрицами $(r_3 - r_1)$, и на соответствующих местах в $(r_2 - r_1)$ находятся нули.

Из справедливости неравенств (5) и (6) следует, что

$$d(R_1, R_3) = d(R_1, R_2) + d(R_2, R_3).$$

5. Отношение „между“. Определение 9. *Признак R находится между признаками R_1 и R_2, если $R_1 \cap R_2 \subseteq R \subseteq R_1 \cup R_2$.*

Это будем обозначать следующим образом: $R \in [R_1, R_2]$.

Определение 10. *Под K-окрестностью признака R будем понимать множество всех признаков $\{R_i, i \in I\}$, для которых $d(R, R_i) = K (K \geq 2)$.*

Окрестность признака R состоит из двух видов признаков:

1. $R_i \subset R$ — будем обозначать их через $\underline{R}_i$,

2. $R_i \supset R$ — будем обозначать их через $\overline{R}_i$.

В 2-окрестности признака T нет признаков вида $\underline{R}$, а для признака U не существуют признаков вида $\overline{R}$, где через T обозначен признак, который разбивает Ω на N одноэлементных множеств, а через U — признак, имеющий одну категорию, совпадающую с Ω.

Теорема 6. *2-окрестность каждого произвольного признака содержит $(N^2-N)/2$ признаков.*

Доказательство. Пусть матрица сояз<ei r признака R содержит P единиц ($P \geqq N$). Тогда в 2-окрестности R будут $(P-N)/2$ признаков вида R_k ($k=1, 2, \ldots, (P-N)/2$), которые получаются удалением из матрицы связей $\bar{r}$ различных пар симметрических единиц и $(N^2-P)/2$ признаков вида $\bar{R}_j$ ($j=1, 2, \ldots, (N^2-P)/2$), полученных после добавления симметрических пар единиц к r. Сумма этих двух видов множеств дает число признаков 2-окрестности R, и она как раз $(N^2-N)/2$.

Определение 11. *Линейным отрезком между признаками R' и R'' будем называть такую последовательность различных признаков $R_1, R_2, \ldots, R_k$, для которой:*

1) $R_1 = R' \cap R''$ и $R_k = R' \cup R''$;

2) $R_i \in [R_m, R_e]$ для $m \leqq i \leqq l$;

3) *Из $R \in [R_i, R_{i+1}]$ следует, что или $R = R_i$, или $R = R_{i+1}$.*

Определение 12. *Два различных признака R_1 и R_2 будем называть соседними, если между ними нет другого признака.*

Очевидно, что линейный отрезок является последовательностью соседних признаков, лежащих между двумя данными. Расстояние между двумя соседними признаками равно 2, поскольку их матрицы связей отличаются на пару симметрических единиц.

От способа образования линейного отрезка следует, что для двух признаков существует $[1/2d(R' \cap R'', R' \cup R'')]!$ различных линейных отрезков.

Теорема 7. *Если два признака R_1 и R_2 — соседние, то для них выполняется или $R_1 \subset R_2$, или $R_2 \subset R_1$.*

Доказательство. Образуем признак $R_1 \cap R_2$. Для него выполняется $R_1 \cap R_2 \in [R_1, R_2]$, но по условию R_1 и R_2 — соседние, следовательно, или $R_1 \cap R_2 = R_2$, или $R_1 \cap R_2 = R_1$. Отсюда получаем, что или $R_1 \subset R_2$, или $R_2 \subset R_1$.

Обозначим через $H(X)$ множества всех номинальных признаков, определенных над множеством Ω, содержащих N объектов. Поскольку $H(X)$ — подпространство пространства $\pi(R)$, то понятия „между“ и „соседние элементы“ справедливы и для него, имея в виду, что объединение двух номинальных признаков в большинстве случаев не принадлежит $H(X)$. В этом пространстве справедлива теорема 5, т. е. если X_1 и X_2 — два соседних номинальных признака, то для них выполняется или $X_1 \subset X_2$, или $X_2 \subset X_1$.

Теорема 8. *Для того чтобы номинальный признак X имел номинальный признак в своей 2-окрестности, необходимо, чтобы он имел двухэлементную категорию или хотя бы две одноэлементные категории.*

Доказательство. Пусть X' — номинальный признак, принадлежащий 2-окрестностям признака X. Следовательно, признаки X и X' будут соседними в ($\pi(R)$ и, по теореме 7, или $X \subset X'$, или $X' \subset X$.

1. Пусть $X \subset X'$, т. е. признак X' получается посредством объединения некоторых из категорий признака X, например, A_P-ая категория (с P-элементами) и A_S-ая категория объединены в одну A'_{P+S} для признака X'. Категории A_P и A_S определяют в матрице связей r признака $X-(P^2+S^2)$ единиц, а категория $A'_{P+S}-(P-S)^2$ единиц в r'. Следовательно, $d(X, X')=(P+S)^2-(P^2+S^2)=2PS$.

Поскольку X' принадлежал 2-окрестности X, выполнено $d(X, X')=2$, откуда $P=S=1$, т. е. категории, которые мы объединяем, должны быть одноэлементными.

2. Пусть $X' \subset X$. В этом случае пусть категория A_K получена посредством объединения категорий A'_m и A'_n признака X' $(k=m+n)$. Тогда $d(X, X')=k^2-(m^2+n^2)=2mn$.

Снова получаем, что должно выполняться $m=n=1$, что означает наличие двухэлементной категории для признака X. Это и доказывает теорему.

Следствие. *Если номинальный признак X имеет P одноэлементных категорий и k двухэлементных, то у него в его 2-окрестности будут* $\left[\binom{P}{2}+k\right]$ *номинальных признаков.*

6. Расположение элементов пространств $\pi(R)$ и $H(X)$.

Определение 13. *Будем говорить, что элементы в одном метрическом пространстве равномерно расположены* (распределены), *если расстояние между любыми двумя соседними элементами равно одному и тому же числу $k \in \mathbf{R}^+$.*

Пространство $\pi(R)$ — пространство с равномерно распределенными элементами, так как если R_1 и R_2 — два его произвольные соседние элемента, то $d(R_1, R_2)=2$.

Теорема 9. *Элементы пространства $H(X)$ неравномерно распределены в нем.*

Доказательство. Возьмем номинальный признак X, у которого нет одноэлементных и двухэлементных категорий. По теореме 8, в 2-окрестности этого признака нет другого номинального признака, т. е. для каждого его соседнего номинального признака X_k(в $H(X)$) выполнено

$$d(X, X_k)>2. \tag{8}$$

С другой стороны, если X' — номинальный признак с хотя бы одной двухэлементной категорией, то, используя теорему 8, можем сказать, что существует хотя бы один номинальный признак X'', принадлежащий 2-окрестности признака X', т. е.

$$d(X', X'')=2. \tag{9}$$

Из (8) и (9) заключаем, что элементы пространства $H(X)$ неравномерно распределены в нем.

Теорема 10. *Число признаков, образующих один линейный отрезок, определенный элементами R' и R'', равняется* $[1/2d(R' \cap R'', R' \cup R'')+1]$.

Доказательство. Возьмем линейный отрезок

$$R' \cap R''=R_1 \subset R_2 \subset R_3 \subset \cdots \subset R_k=R' \cup R''. \tag{10}$$

Используя теорему 5, можем записать

$$d(R_1, R_k)=d(R_1, R_2)+d(R_1, R_3)+\cdots+d(R_{k-1}, R_k).$$

Поскольку, по определению, для любых двух последовательных значений j и $(j+1)$ индексов признаки соседние, то $d(R_j, R_{j+1})=2$, и, следовательно, $d(R_1, R_k)=2(k-1)$. Из этого равенства можем определить K, что даст число признаков на линейном отрезке (10): $K=1/2\, d(R' \cap R'', R' \cup R'')+1$, чем теорема доказана.

Возьмем два произвольных признака R' и R'' и образуем последовательности:

(11) $$R' \cap R'' = R_1 \subset R_2 \subset R_3 \subset \cdots \subset R_{k-1} \subset R_k = R'';$$

(12) $$R' = R^1 \subset R^2 \subset R^3 \subset \cdots \subset R^{s-1} \subset R^s = R' \cup R'';$$

(13) $$R' \cap R'' = {}_1R \subset {}_2R \subset {}_3R \subset \cdots \subset {}_{p-1}R \subset {}_pR = R';$$

(14) $$R'' = {}^1R \subset {}^2R \subset {}^3R \subset \cdots \subset {}^{m-1}R \subset {}^mR = R' \cup R'',$$

причем для каждой из них и для любых двух последовательных индексов j и $(j+1)$ соответствующие признаки — соседние.

Так, определенные последовательности признаков имеют следующие свойства:

Лемма 1. *Последовательности* (11) *и* (12) *имеют одно и то же число элементов.*

Доказательство. Используя определение расстояния между двумя признаками (4), записываем:

$$d(R' \cap R'', R'') = d(R_1, R'') = |r_1 \triangle r''| - N.$$

Из $R' \cap R'' \subset R''$ получаем $r_1 \subset r''$. Следовательно, $r_1 \triangle r'' = (r_1 \cup r'') - (r_1 \cap r'') = r'' - r_1 = r'' - r'$ и

(15) $$d(R' \cap R'', R'') = |r'' - r'| - N.$$

Аналогично получаем

(16) $$d(R', R' \cup R'') = |r'' - r'| - N,$$

используя, что $r' \subset r^s$, так как $R' \subset R' \cup R''$.

Из равенств (15) и (16) записываем

(17) $$d(R' \cap R'', R'') = d(R', R' \cup R'') = |r'' - r'| - N.$$

Используя теорему 10 и равенство (17), заключаем, что у последовательностей (11) и (12) одинаковое число элементов и, следовательно, $k = s$.

Лемма 2. *У последовательностей* (13) *и* (15) *одинаковое число элементов.*

Доказательство проходит аналогично предыдущему. Получаем

(18) $$d(R' \cap R'', R') = d(R'', R' \cup R'') = |r' - r''| - N.$$

Теорема 11. *Для элементов последовательностей* (11), (12), (13) *и* (14) *выполняются неравенства*

(19) $$d(R_\mu, R^\mu) \geq d(R', R' \cap R'') \quad (\mu = 1, 2, \ldots, k);$$

(20) $$d({}_\nu R, {}^\nu R) \geq d(R' \cap R'', R'') \quad (\nu = 1, 2, \ldots, P).$$

Доказательство. Справедливо следующее включение:

(21) $$(r_1 \cup r^1) - (r_1 \cap r^1) \subseteq (r_\mu \cup r^\mu) - (r_\mu \cap r^\mu).$$

Доказательство проведем следующим образом: пусть $r_{ij} = 1$ — произвольный ненулевой элемент в левой стороне включения (21). Это означает, что $(r_{1,ij} = 1 \wedge r^1_{ij} = 0) \vee (r_{1,ij} = 0 \wedge r^1_{ij} = 1)$, где через $r_{1,ij}$ обозначены элементы матрицы r_1. Случай $(r_{1,ij} = 1 \wedge r^1_{ij} = 0)$ невозможен, поскольку $r_1 \subset r^1$.

Из $r_{1,ij}=0$, $r^1_{ij}=1$ и $r'\cap r''\subset r''$ следует равенство $r''_{ij}=0$, так как при допущении, что $r''_{ij}=1$, будет следовать, что $r_{1,ij}=1$, так как $r^1_{ij}=1$. С другой стороны, $R_\mu\subset R''$ и, следовательно, $r_{\mu,ij}=0$, откуда получаем, что ij-ый элемент $r_\mu\cap r^\mu$ — ноль. Из включения $r^1\subset r^\mu$ следует, что $r^\mu_{ij}=1$ и, следовательно, ij-ый элемент матрицы в правой стороне включения (21) равняется единице, что показывает, что оно справедливо, и можем записать его в виде $|r_1\triangle r^1|\leqq|r_\mu\triangle r^\mu|$. Но это дает расстояние между соответствующими признаками, т. е. $d(R_1, R^1)\leqq d(R_\mu, R^\mu)$.

Аналогично доказываются и неравенства (10).

Теорема 12. *Если признак R принадлежит отрезку $[R', R'']$, то $d(R', R'')=d(R', R)+d(R, R'')$.*

Доказательство. Возьмем сначала вместо произвольного признака R отрезка признак $R'\cap R''$.

Из (17) и (18) имеем $d(R', R'\cap R'')=|r'-r''|-N$, $d(R'\cap R'', R'')=|r''-r'|-N$. Используя определение разности признаков, получим, что $|r'-r''|+|r''-r'|-2N$ даст как раз число несовпадений единиц матриц r' и r''. Следовательно,

$$d(R', R'')=d(R', R'\cap R'')+d(R'\cap R'', R''). \tag{22}$$

Пусть теперь R — произвольный признак отрезка $[R', R'']$. Тогда он принадлежит некоторой из четырех последовательностей (11), (12), (13) и (14). Допускаем, что R принадлежит последовательности (13), т. е. $R'\cap R''\subset R\subset R'$. Тогда по теореме 5 можем записать для равенства (22)

$$d(R', R'')=d(R', R)+d(R, R'\cap R'')+d(R'\cap R'', R''). \tag{23}$$

Используя аксиому 3, для введенного в $\pi(R)$ расстояния получаем, что

$$d(R, R'\cap R'')+d(R'\cap R'', R'')\geqq d(R, R''). \tag{24}$$

Подставляя (24) в (23), находим

$$d(R', R'')\geqq d(R', R)+d(R, R''). \tag{25}$$

Применяя снова аксиому 3, для признаков R', R и R'' получаем неравенство

$$d(R', R'')\leqq d(R', R)+d(R, R''). \tag{26}$$

Из (25) и (26) следует непосредственно равенство $d(R', R'')=d(R', R)+d(R, R'')$.

Следствие. *Пусть $R_1, R_2, \ldots, R_k$ — линейный отрезок для признаков R' и R''. Тогда*

$$d(R', R'')=\sum_{i=1}^{k-1} d(R_i, R_{i+1}).$$

7. Выпуклые множества и выпуклые оболочки. Наличие понятия „между“ в пространстве $\pi(R)$ позволяет естественным образом ввести понятие „выпуклое множество“ в нем.

Определение 14. *Множество D элементов $\pi(R)$ называется выпуклым, если вместе с любыми двумя его элементами оно содержит все элементы $\pi(R)$, лежащие между ними.*

Используя введенные выше обозначения, можем сказать, что множество D — выпуклое, если из $R', R''\in D$ и $R\in[R', R'']$ следует, что и $R\in D$.

Определение 15. *Будем говорить, что признак R' минимален в множестве $P \subseteq \pi(R)$, если для каждого признака R множества P выполняется $|r'| \leq |r|$.*

Теорема 13. *В выпуклом множестве $P \subseteq \pi(R)$ есть единственный минимальный элемент $R_{\min}$ и максимальный $R_{\max}$.*р

Доказательство. Допустим, что есть два азличных элемента R' и R'' в множестве P. Образуем признак $R' \cap R''$. Он принадлежит также и P, поскольку $R' \cap R'' \in [R', R'']$. С другой стороны, $|r' \cap r''| < r'| = |r''|$, поскольку R' и R'' — различные. Следовательно, $R' \cap R''$ — минимальный в P, а R' и R'' — неминимальные. Приходим к противоречию.

Аналогично доказывается и существование единственного максимального элемента.

Теорема 14. *Для любых двух соседних номинальных признаков X' и X'' (соседние в пространстве $H(X)$) существует хотя бы одна последовательность из $[1/2\, d(X', X'')-1]$ соседних аддитивных признаков, для которых выполнено $X' \subset L_1 \subset L_2 \subset \cdots \subset L_k \subset X''$.*

Замечание. Надо иметь в виду, что для номинальных признаков, которые соседние в пространстве $\pi(R)$ (см. теорему 8), не существует такой последовательности, потому что между ними нет никаких признаков из $\pi(R)$.

Доказательство. Имея в виду теорему 7, можем записать, что $X' \subset X''$. Следовательно, линейный отрезок, определенный признаками X' и X'', может быть представлен так: $X' \subset R_1 \subset R_2 \subset \cdots \subset R_S \subset X''$. Поскольку X' и X'' — соседние в $H(X)$, то между признаками R_i, $i=1, 2, \ldots, s$ не может быть номинальных.

По теореме 10, число признаков на одном отрезке равняется $[1/2\, d(X', X'')+1]$. Исключая из общего числа признаков на отрезке оба номинальных признака, получаем $[1/2\, d(X', X'')-1]$ аддитивных признаков.

В заключение нам хотелось бы отметить, что:

1) Полученное пространство неупорядоченных качественных признаков номинального и аддитивного типа замкнуто по отношению введенных в него операций.

2) Элементы пространства $\pi(R)$ равномерно расположены в нем относительно введенной метрики, рассматриваемой в [6; 3; 7] и др.

3) Полученные результаты дают возможность для аддитивной аппроксимации номинальных признаков.

ЛИТЕРАТУРА

1. А. Н. Колмогоров, С. В. Фомин. Элементы теории функций и функционального анализа. Москва, 1976.
2. Б. Г. Миркин, Анализ качественных признаков и структур. Москва, 1980.
3. Б. Г. Миркин, Л. Б. Черный. Об измерении близости между различными разбиениями конечного множества объектов. *Автоматика и телемеханика*, 5, 1970.
4. В. Здравков, В. Тасова. Понятието зависимост в случая на нечислови признаци и начин за неговото измерване. *Социално управление*, 5, 1980, 15—24.
5. В. Б. Кузьмин. Построение групповых решений в пространствах четких и нечетких бинарных отношений. Москва, 1982.
6. J. Kemeny, J. Snell. Mathematical Models in Social Sciences. Providance, 1962.
7. K. Bogarth. Distance in a set of transitive relation. *J. Math. Soc.*, 2, 1972.

Поступила 6. 9. 1983 *В переработанном виде 12. 2. 1984*

Пловдив 4000, ул. Цариградско шосе № 39