Часть I

Информационный подход в распознавании образов

Разделы

Информация по Хартли: введение

Определение (информация по Хартли)

Если Ω — конечное множество, то *информация* $I(\omega)$ *его элемента* $\omega \in \Omega$ есть величина

$$I(\omega) = \log|\Omega|.$$

Если на Ω задано отношение эквивалентности $\sim$, то *информация* $I_\sim(\omega)$ *элемента* ω *по эквивалентности* ω есть

$$I_\sim(\omega) = \log|[\omega]_\sim|$$

(т.е. $I_\sim(\omega)$ есть логарифм мощности класса эквивалентности, в который попадает элемент ω).

Здесь и далее все логарифмы — по основанию 2 (информация измеряется в битах).

Р. Хартли

Ральф Хартли
(Ralph Vinton Lyon Hartley, 1888–1970) — американский исследователь в области электроники (ему принадлежат более 70 патентов); ввёл в 1928 г. логарифмическую меру информации, которая называется *хартлиевским количеством информации*.

$$I = \log_2 N \qquad \boxed{N = 2^I}$$

I – количество информации;
N – количество возможных событий;
Все события – равновероятны.

Эквивалентность на множестве и информация

Утверждение

Если $N = |\Omega/\!\sim|$ *— число классов эквивалентности, то*

$$\sum_{\omega\in\Omega} 2^{-(I_\sim(\omega)+\log N)} = 1.$$

Доказательство

Пусть W_i *—* i*-й класс эквивалентности.*
Тогда (суммирование «по элементам» $\to$ *«по классам эквивалентности»)*

$$\sum_{\omega\in\Omega} 2^{-(I_\sim(\omega)+\log N)} = \frac{1}{N}\sum_{i=1}^{N} |W_i|\, 2^{-\log|W_i|} = \frac{1}{N}\sum_{i=1}^{N} 1 = 1.$$

Неравномерные коды

Пусть

- $|X|$ — конечный алфавит;
- X^n — множество всех слов длины n в X;
- $B = \{0, 1\}$ — двоичный алфавит;
- B^+ — множество всех слов длины > 0 в B.

Определение

Неравномерным кодом (*сжимающим отображением*) называют отображение

$$\psi : X^n \to B^+,$$

определённое для каждого $n = 1, 2, \ldots$.

Код называется *префексным* или *дешифрируемым*, если $\psi(x)$ не является началом никакого префексного слова $\psi(x'), x' \neq x$.

Длина кодового слова $\psi(x)$ обозначается $l(x)$.

Неравенство Крафта-Макмиллана —

— критерий существования существования префиксных кодов, обладающих заданным набором длин кодовых слов.

Утверждение (неравенство Крафта)
Для существования префиксного кода C *из* N *слов с длинами* $l_1, \ldots, l_N$ *необходимо и достаточно, чтобы*

$$\sum_{i=1}^{N} 2^{-l_i} \leqslant 1.$$

Пример: $C = \{01, 10, 110, 001\}$, $l_1 = l_2 = 2$, $l_3 = l_4 = 3$,

$$\frac{1}{2^2} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^3} = \frac{6}{8} = \frac{3}{4} < 1.$$

Неравенство Крафта...

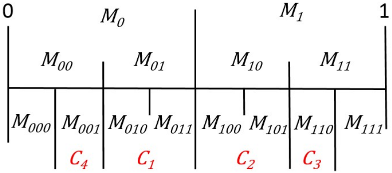
Доказательство

Рассмотрим код C из I слов с длинами $l_1, l_2, \ldots, l_I$ и отрезок $[0;1]$ на числовой прямой.

1. *Разделим его пополам, левую половину обозначим M_0, а правую — M_1.*
2. *Поделим M_0 пополам и обозначим его левую половину M_{00}, а правую M_{01}, и, проделав то же самое с M_1, получим M_{10}, а левую M_{11}... и т.д., пока длина индекса полученного отрезка M_j не станет $= \max\{l_1, l_2, \ldots, l_I\}$.*

В нашем примере

$$C = \{01, 10, 110, 001\}$$



Неравенство Крафта...

Ясно, что:

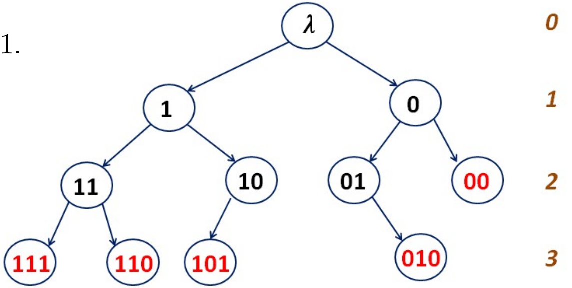
- *любому кодовому слову C_j сопоставлен свой отрезок M_{C_j} — например, кодовому слову 110 соответствует отрезок M_{110};*
- *длина отрезка M_{C_i} равна 2^{-l_i};*
- *если код C — префиксный, то никакие два отрезка не пересекаются;*
- *сумма длин отрезков не превзойдет 1, то есть*
$$\sum_{i=1}^{N} |M_{C_i}| = \sum_{i=1}^{N} 2^{-l_i} \leqslant 1.$$

Следствие: существует префиксный двоичный код с длинами
$$l(\omega) = \lceil I(\omega) + \log N \rceil.$$

Неравенство Крафта: пример

Укоренённое двоичное дерево можно рассматривать как графическое описание префиксного кода над B.

$\frac{1}{4} + 4\left(\frac{1}{8}\right) = \frac{3}{4} \leqslant 1.$



Неравенство Крафта для таких деревьев: $\sum_{x \in L} 2^{-\mathrm{depth}(x)} \leqslant 1$, где L — множество листьев дерева, а $\mathrm{depth}(x)$ — глубина листа x (число рёбер на пути от корня λ до x).

Разделы

Действие группы на множестве: два определения

- Группа $\mathcal{G} = \langle\, G, \circ, e \,\rangle$, $|G| = n$.
- Множество Ω, $|\Omega| = N$.
  - $Bij(\Omega)$ — множество всех биекций (перестановок) элементов Ω.
  - $Symm(\Omega)$ — симметрическая группа множества Ω:

$$Symm(\Omega) = \langle\, Bij(\Omega), *, 1_\Omega \,\rangle = S_N$$

Определение (I)

$$\alpha \in \mathrm{Hom}\,(\,\mathcal{G}, Symm(\Omega)\,).$$

Обозначение действия α группы $\mathcal{G}$ на множестве Ω — $\mathcal{G} \underset{\alpha}{:} \Omega$.

Пример: $\forall g \in G : \alpha(g) = 1_\Omega$ — тривиальное действие.

Действие группы на множестве: два определения...

Определение (II)

$$\alpha = \langle G, \Omega; \circ, *, e \rangle ,$$

где

$$G \times G \xrightarrow{\circ} G \text{ — групповая операция;}$$
$$G \times \Omega \xrightarrow{*} \Omega \text{ — новая операция.}$$

Аксиомы для операций:

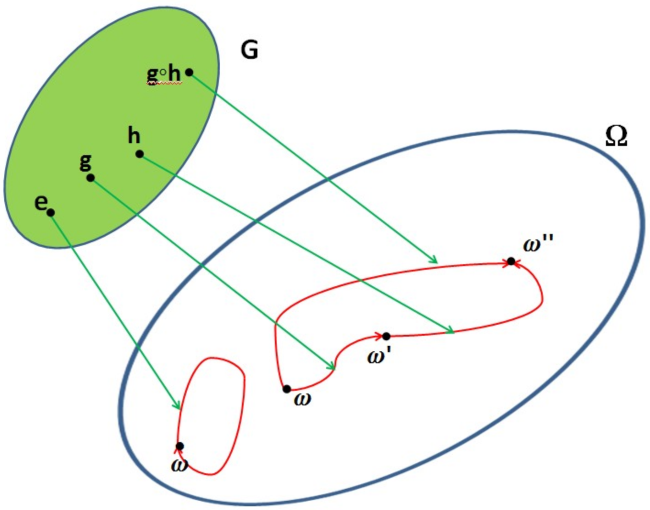
$$1).\ e * \omega = \omega; \qquad 2.\ (g \circ h) * \omega = h * (g * \omega).$$

Запись операции $*$: $g * \omega = g(\omega) = \omega'$.

Аксиомы: $e(\omega) = \omega$ и $(g \circ h)(\omega) = h(g(\omega))$.

Т.е. элементы g группы G порождают перестановки на Ω, обладающие вышеуказанными свойствами.

Действие группы на множестве: схема



Для данной перестановки g:

Введём отношение эквивалентности $\sim_g$ на Ω —

$$\omega \sim_g \omega' \Leftrightarrow \exists\, k \in \mathbb{Z} : g^k(\omega) = \omega'$$

Смежные классы эквивалентности $\sim_g$ — *g-циклы*.

Обозначения:

- $C(g)$ — число g-циклы (смежных классов по эквивалентности $\sim_g$).
- $\nu_1, \nu_2, \ldots, \nu_N$ (или $\nu_1(g), \nu_2(g), \ldots, \nu_N(g)$) — количества циклов длины $1, 2, \ldots, N$.
- $\langle\, \nu_1, \nu_2, \ldots, \nu_N \,\rangle = Type(g)$ — *тип перестановки* g (упорядоченная совокупность количеств циклов).

Понятно, что $C(g) = \sum\limits_{k=1}^{N} \nu_k(g)$ и $\sum\limits_{k=1}^{N} k \cdot \nu_k(g) = N$.

Тип перестановки: пример

Пусть

$$\Omega = \{1, \ldots, 10\},$$
$$g = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 9 & 6 & 1 & 8 & 5 & 2 & 7 & 10 & 3 & 4 \end{pmatrix} =$$
$$= (1,9,3)(2,6)(4,8,10)(5)(7) = (1,9,3)(2,6)(4,8,10)$$

Тогда

$$Type(g) = \langle 2, 1, 2, 0, \ldots, 0 \rangle,$$

$$C(g) = 2+1+2 = 5, \qquad |\Omega| = 2\cdot 1 + 1\cdot 2 + 2\cdot 3 = 10.$$

По всей группе $\mathcal{G}$:

Отношение эквивалентности $\sim_{\mathcal{G}}$ на Ω —

$$\omega \sim_{\mathcal{G}} \omega' \stackrel{\text{def}}{=} \underset{G}{\exists}\, g : g(\omega) = \omega'.$$

- Классы этой эквивалентности называют *орбитами*.
- Класс эквивалентности, в которую попадает элемент ω обозначают $\mathrm{Orb}\,(\omega)$.
- Число орбит — $C(\mathcal{G})$.
- Орбиты образуют разбиение множества Ω.

Если $C(\mathcal{G}) = 1$ (любой элемент Ω может быть переведён в любой), то действие $\mathcal{G} \underset{\alpha}{:} \Omega$ называют *транзитивным*.

Соответствующую эквивалентность будем обозначать так же, как и порождающее её действие группы на множестве — α.

Очевидно $\quad I_\alpha(\omega) = \log|\mathrm{Orb}\,(\omega)|.$

Фиксатор и стабилизатор. Лемма Бёрнсайда

Будем решать уравнение $g(\omega) = \omega$.

При выполнении этого равенства можно фиксировать ω или g.

1. Фиксируем g, т.е. находим все элементы множества Ω, которые перестановка g оставляет на месте — *фиксатор перестановки* $g \in \mathcal{G}$:

$$\{\, \omega \in \Omega \mid g(\omega) = \omega \,\} \stackrel{\text{def}}{=} \operatorname{Fix}(g) \subseteq \Omega.$$

2. Фиксируем ω, т.е. находим все перестановки g, которые оставляют данный элемент неподвижным *стабилизатор элемента* $\omega \in \Omega$:

$$\{\, g \in G \mid g(\omega) = \omega \,\} \stackrel{\text{def}}{=} \operatorname{Stab}(\omega) \subseteq G.$$

Стабилизатор есть подгруппа группы $\mathcal{G}$

Утверждение

$$\mathrm{Stab}\,(\omega) \leqslant \mathcal{G}.$$

Доказательство

Зафиксируем $\omega \in \Omega$ *и рассмотрим* $g, h \in \mathrm{Stab}\,(\omega)$. *Тогда* $g(\omega) = h(\omega) = \omega$ *и* $h^{-1}(\omega) = \omega$. *Следовательно,*

$$(g \circ h^{-1}) * \omega = \omega \Rightarrow g \circ h^{-1} \in \mathrm{Stab}\,(\omega)\,.$$

$\left|\mathrm{Stab}\,(\omega)\right| \geqslant 1$, поскольку всегда $e \in \mathrm{Stab}\,(\omega)$.

Стабилизатор ещё называют *стационарной подгруппой* элемента ω и обозначают G_ω (G_ω не обязательно нормальная подгруппа G).

Элемент множества: длина орбиты и стабилизатор

Лемма

Длина орбиты $\mathrm{Orb}\,(\omega)$ *равна индексу подгруппы* $\mathrm{Stab}\,(\omega)$ *в группе* $\mathcal{G}$*:*

$$|\mathrm{Orb}\,(\omega)| = \frac{|G|}{|\mathrm{Stab}\,(\omega)|} = [G : \mathrm{Stab}\,(\omega)] = [G : G_\omega]\,.$$

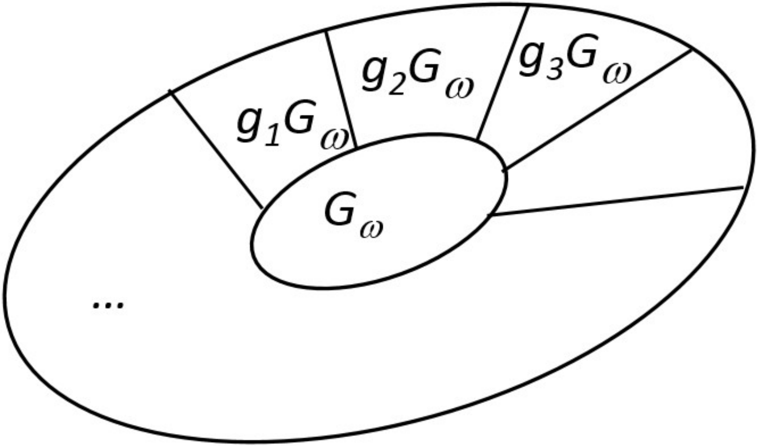
Доказательство

Теорема Лагранжа: $H \leqslant G \Rightarrow |G| = |H| \cdot [G : H]$*,*
где $[G : H]$ *— индекс подгруппы* H *группы* G *= количество классов смежности, и левых, и правых.*

Напоминание: $gH = \{gh \mid h \in H\}$*,* $Hg = \{hg \mid h \in H\}$*.*
Любой представитель данного смежного класса R *переводит* $\omega \to \omega' = \omega'(R)$*:*

$$(g \circ h) * \omega = h * (g * \omega) = g * \omega = g(\omega).$$

Длина орбиты и стабилизатор: иллюстрация



Лемма не-Бёрнсайда

Лемма

$$C(\mathcal{G}) = \frac{1}{|G|}\sum_{g\in G}\Big|\operatorname{Fix}(g)\Big| = \frac{1}{|G|}\sum_{\omega\in\Omega}\Big|\operatorname{Stab}(\omega)\Big|;$$

первое равенство называется леммой Бёрнсайда.

Доказательство

1. Подсчитаем двумя различными способами мощность множества $M = \{\,(g,\omega)\in G\times\Omega \mid g(\omega)=\omega\,\}$*:*

$$\sum_{g\in G}\Big|\operatorname{Fix}(g)\Big| = |M| = \sum_{\omega\in\Omega}\Big|\operatorname{Stab}(\omega)\Big|$$

2. Ясно, что если $\operatorname{Orb}(x)=\operatorname{Orb}(y)$ *(элементы* $x,y\in\Omega$ *принадлежат одному классу эквивалентности по* $\sim_{\mathcal{G}}$*), то*

$$|\operatorname{Stab}(x)| = |G|/|\operatorname{Orb}(x)| = |G|/|\operatorname{Orb}(y)| = |\operatorname{Stab}(y)|\,.$$

Лемма Бёрнсайда...

3. Выберем по представителю $\omega_1, \ldots, \omega_{C(\mathcal{G})}$ *из всех* $C(\mathcal{G})$ *орбит.*
Тогда

$$|M| = \sum_{g \in G} \Big| \mathrm{Fix}\,(g) \Big| = \sum_{i=1}^{C(\mathcal{G})} |\mathrm{Orb}\,(\omega_i)| \cdot |\mathrm{Stab}\,(\omega_i)| =$$

$$= \sum_{i=1}^{C(\mathcal{G})} \frac{|G|}{|\mathrm{Stab}\,(\omega_i)|} \cdot |\mathrm{Stab}\,(\omega_i)| = |G| \cdot C(\mathcal{G}).$$

У. Бёрнсайд

Уильям Бёрнсайд
(William Burnside, 1852–1927)
— английский математик-алгебраист.
«Написал первый трактат о группах на английском языке и был первым, кто разработал теорию групп с современной абстрактной точки зрения».

Также знаменит формулированием *проблемы Бёрнсайда* (1902):
Будет ли конечнопорождённая группа, в которой каждый элемент имеет конечный порядок, обязательно конечной?

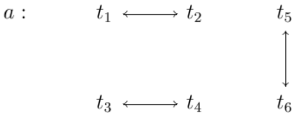
Ответ отрицательный (1992).

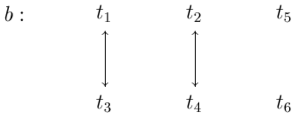
Действие группы на множестве: пример

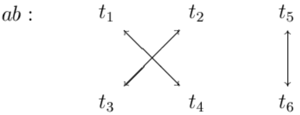
Действие α группы V_4 на множестве $\Omega = \{\omega_1, \ldots, \omega_6\}$

$\circ$	e	a	b	ab
e	e	a	b	ab
a	a	e	ab	b
b	b	ab	e	a
ab	ab	b	a	e

$g * t$	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
e	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
a	ω_2	ω_1	ω_4	ω_3	ω_6	ω_5
b	ω_3	ω_4	ω_1	ω_2	ω_5	ω_6
ab	ω_4	ω_3	ω_2	ω_1	ω_6	ω_5







$$I_\alpha(\omega_1) = \log 4 = 2$$

Действие группы на множестве: пример...

$$Type(e) = \langle\, 6,\, 0,\, 0,\, 0,\, 0,\, 0 \,\rangle\,, \qquad Type(a) = \langle\, 0,\, 3,\, 0,\, 0,\, 0,\, 0 \,\rangle\,,$$
$$Type(b) = \langle\, 2,\, 2,\, 0,\, 0,\, 0,\, 0 \,\rangle\,, \qquad Type(ab) = \langle\, 0,\, 3,\, 0,\, 0,\, 0,\, 0 \,\rangle\,.$$
$$C(e) = 6\,,\; C(a) = C(ab) = 3\,,\; C(b) = 4\,.$$
$$\mathrm{Stab}\,(\omega_1) = \mathrm{Stab}\,(\omega_2) = \mathrm{Stab}\,(\omega_3) = \mathrm{Stab}\,(\omega_4) = e \leqslant V_4\,,$$
$$\mathrm{Stab}\,(\omega_5) = \mathrm{Stab}\,(\omega_6) = \langle\, e,\, b \,\rangle \leqslant V_4\,.$$
$$\mathrm{Fix}\,(a) = \mathrm{Fix}\,(ab) = \varnothing\,, \quad \mathrm{Fix}\,(b) = \{\, \omega_5,\, \omega_6 \,\}\,, \quad \mathrm{Fix}\,(e) = \Omega\,.$$
$$\big|\,\mathrm{Orb}\,(\omega_1)\big| = \frac{4}{1} = 4\,, \quad \big|\,\mathrm{Orb}\,(\omega_5)\big| = \frac{4}{2} = 2\,.$$
$$\frac{1}{4}\sum_{g\in G}\big|\,\mathrm{Fix}\,(g)\big| = \frac{6+2}{4} = 2\,,$$
$$\frac{1}{4}\sum_{t\in\Omega}\big|\,\mathrm{Stab}\,(t)\big| = \frac{4\cdot 1 + 2\cdot 2}{4} = 2\,.$$

Разделы

Композиция слова

Пусть

- $X = \{ a_1, \ldots, a_q \}$ — конечный алфавит,
- $\Omega = X^n$ — множество всех слов длины n в алфавите X,
- на Ω действует симметрическая группа S_n, переставляющая буквы в словах.

Определение

Композицией слова $\omega \in X^n$ называется кортеж

$$(m_1, \ldots, m_q),$$

где m_i — число вхождений буквы a_i в слово ω, $i = \overline{1, q}$.

Перестановка букв в слове не меняет количеств их вхождений $\Leftrightarrow$ и поэтому все слова одной орбиты имеют одну и ту же композицию.

Орбита слова: пример

Пример

$X = \{a, b\}$, $n = 5$, $\Omega = \{aaaaa, aaaab, \ldots, bbbbb\} = X^5$,
$|\Omega| = 2^5 = 32 = N$

Действие $S_5 \underset{\alpha}{:} \Omega$ симметрической группы S_5 всех $5! = 120$ перестановок на слова из Ω.

Пусть $\omega = ababb$, $\operatorname{Orb}(ababb) = ?$

$$|\operatorname{Orb}(ababb)| = [\, S_5 : \operatorname{Stab}(ababb)\,] = \frac{|S_5|}{|\operatorname{Stab}(ababb)|},$$

$$\operatorname{Stab}(ababb) = \{\, e, (13), (245), (254), (54), \ldots, (13)(245), \cdots \}.$$

$$\operatorname{Stab}(ababb) \cong \mathbb{Z}_2 \times S_3 \Rightarrow |\operatorname{Stab}(ababb)| = 2 \cdot 6 = 12 \Rightarrow$$

$$\Rightarrow |\operatorname{Orb}(ababb)| = \frac{120}{12} = 10$$

Орбита слова: пример...

$\mathrm{Orb}\,(ababb)$:

$$
\begin{array}{ccccc}
aabbb & ababb & abbab & abbba & baabb \\
babab & babba & bbaab & bbaba & bbbaa
\end{array}
$$

Поскольку все слова одной орбиты имеют одну и ту же композицию, то

$$|\mathrm{Orb}\,(ababb)| = \frac{5!}{2!\cdot 3!} = 10$$

0-информация слова

Определение

0-*информацией* $I_0(x)$ слова $x \in X^n$ называется величина

$$I_0(x) = I_\alpha(x) = \log|\mathrm{Orb}\,(x)| = \log\frac{n!}{m_1! \cdot \ldots \cdot m_q!},$$

где $\mathrm{Orb}\,(x)$ — орбита слова x, при действии $S_n \underset{\alpha}{:} X^n$.

Пример

$$I_0(abaa) = \log\frac{4!}{3!\,1!} = 2,\ I_0(abab) = \log\frac{4!}{2!\,2!} = \log 6 \approx 3{,}32$$

$I_0(aaaa) = 0,\ I_0(abcd) = \log 24 \approx 4{,}58$ — минимальное и максимальное значение $I_0(\cdot)$для слов из 4 элементов

Меньше симметрий в слове $\Rightarrow$ больше его 0-информация. Наличие симметрий позволяет эффективно кодировать (сжимать) слово.

Условная 0-информация слова

Вместе с алфавитом X будем рассматривать алфавит Y (не исключено $X = Y$) и множество слов Y^n.

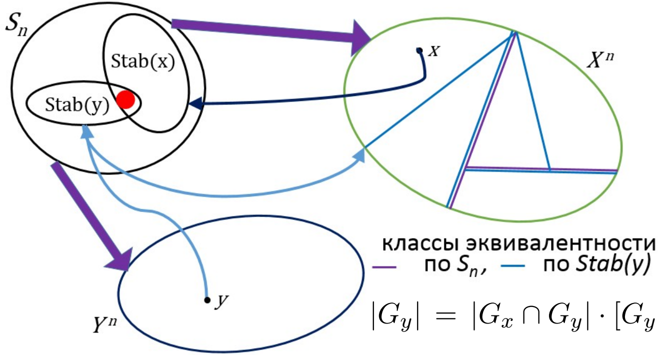
$$\begin{array}{ccc} x = & (x_1, \ldots, & x_n) \\ \Updownarrow & \updownarrow & \updownarrow \\ y = & (y_1, \ldots, & y_n) \end{array}$$

Группа симметрическая группа S_n действует на X и Y.

Рассмотрим стационарная подгруппу $\mathrm{Stab}\,(y) \leqslant S_n$ слова $y \in Y^n$.

Под действием $\mathrm{Stab}\,(y)$ слова из X^n разбиваются на орбиты, которые назовём *условными*.

Условная 0-информация слова...



$$|G_y| = |G_x \cap G_y| \cdot [G_y : (G_x \cap G_y)]$$

Определение

Условной 0-информацией $I_0(x/y)$ слова $x \in X^n$ относительно слова $y \in Y^n$ называют величину

$$I_0(x/y) = \log [\, G_y : (G_x \cap G_y) \,] = \log \frac{|\mathrm{Stab}\,(y)|}{|\mathrm{Stab}\,(x) \cap \mathrm{Stab}\,(y)|}.$$

Условная 0-информация: пример

$$I_0(x/y) \;=\; \log\,[\,G_y : (G_x \cap G_y)\,] \;=\; \log\,\frac{|\mathrm{Stab}\,(y)|}{|\mathrm{Stab}\,(x) \cap \mathrm{Stab}\,(y)|}.$$

Пусть $X \;=\; \{\,a,\,b\,\},\; Y \;=\; \{\,0,\,1\,\}$.

1. $\underline{n = 3,\; x = aab,\; y = 111}$:

$$I_0(x) = \log\tfrac{3!}{2!1!} = \log 3,\;\; I_0(y) = \log\tfrac{3!}{3!} = \log 1 = 0,$$
$$\mathrm{Stab}\,(x) \;=\; \{\,e,\,(12)\,\},\; \mathrm{Stab}\,(y) \;=\; S_3,\;\; |S_3| = 6,$$
$$\mathrm{Stab}\,(x) \cap \mathrm{Stab}\,(y) = \mathrm{Stab}\,(x) \;\; \text{и}$$
$$I_0(x/y) \;=\; \log\,\tfrac{6}{2} = \;\log 3,\; I_0(y/x) \;=\; \log\tfrac{2}{2} \;=\; 0.$$

Условная 0-информация: пример

$$I_0(x/y) \;=\; \log\left[\, G_y : (G_x \cap G_y)\,\right] \;=\; \log\; \frac{|\mathrm{Stab}\,(y)|}{|\mathrm{Stab}\,(x) \cap \mathrm{Stab}\,(y)|}.$$

2. $\underline{n = 5,\; x = aabab,\; y = 11000}$:

$$I_0(x) \;=\; I_0(y) \;=\; \log\frac{5!}{2!\cdot 3!} \;=\; \log 10$$

$$\mathrm{Stab}\,(x) \;=\; \{\, e,\, (124),\, (142),\, (12),\, (24),\, (14),\, (35),\, \ldots \} \;\cong$$

$$\cong\; \mathbb{Z}_2 \times S_3 \;\Rightarrow\; |\,\mathrm{Stab}\,(x)| \;=\; 12$$

$$\mathrm{Stab}\,(y) \;=\; \{\, e,\, (12),\, (345),\, (354),\, (34),\, (45),\, (35)\,\} \;\cong$$

$$\cong\; \mathbb{Z}_2 \times S_3 \;\Rightarrow\; |\,\mathrm{Stab}\,(y)| \;=\; 12$$

$$\mathrm{Stab}\,(x) \cap \mathrm{Stab}\,(y) \;=\; \{\, e,\, (12),\, (35),\, (12)(35)\,\} \;\Rightarrow$$

$$\Rightarrow\; |\,\mathrm{Stab}\,(x) \cap \mathrm{Stab}\,(y)| \;=\; 4$$

$$I_0(x/y) \;=\; I_0(y/x) \;=\; \log\frac{12}{4} \;=\; \log 3$$

Задание слов таблицами: пример

Слово $x = (abaacbca) \in \{ a, b \}^8$ можно задать в виде таблицы номеров вхождений его символов

$$x = \left| \begin{array}{cccc} 1 & 3 & 4 & 8 \\ 2 & 6 & & \\ 5 & 7 & & \end{array} \right| \qquad \begin{array}{c} (a) \\ (b) \\ (c) \end{array}$$

по которой оно однозначно восстанавливается с точностью до переименований букв.

1. На позиции 1, 3, 4 и 8 устанавливаем некоторую букву, например d:

1	2	3	4	5	6	7	8
d		d	d				d

Задание слов таблицами: пример

2. Далее, на позициях 2 и 6 устанавливаем некоторую другую букву, например a:

1	2	3	4	5	6	7	8
d	a	d	d		a		d

3. Наконец, оставшиеся позиции занимаем третьей буквой, например b:

1	2	3	4	5	6	7	8
d	a	d	d	b	a	b	d

В результате получаем слово y, полученное из исходного слова x переименованием букв

$$a \mapsto d,\ b \mapsto a,\ c \mapsto b$$

(в криптографии такой приём называется *шифром Цезаря*).

Произведение слов

Утверждение

$$I_0(x/y) = 0 \Leftrightarrow G_y \subseteq G_x$$

Доказательство

$$I_0(x/y) = \log \frac{|G_y|}{|G_x \cap G_y|} = 0 \Leftrightarrow G_y = G_x \cap G_y \Leftrightarrow G_y \subseteq G_x.$$

Если $I_0(y/x) = I_0(x/y) = 0$, то слова назовём *эквивалентными*.

Произведение слов

Пусть $x = (a_1, \ldots, a_n) \in X^n$, а $y = (b_1, \ldots, b_n) \in Y^n$, тогда

$$z = x \times y = (\,(a_1, b_1), \ldots, (a_n, b_n)\,)$$

— *произведение слов* x *и* y, слово в алфавите $Z = X^n \times Y^n$.

Утверждение

$$I_0(x \times y) = I_0(x) + I_0(y/x) = I_0(y) + I_0(x/y) = I_0(y \times x)$$

Доказательство

Очевидно $\mathrm{Stab}\,(x \times y) = \mathrm{Stab}\,(x) \cap \mathrm{Stab}\,(y)$, *или, упрощая обозначения* $G_{x \times y} = G_x \cap G_y$. *Далее по теореме Лагранжа:*

$$[\,G : (G_x \cap G_y)\,] = [\,G : G_x\,] \cdot [\,G_x : (G_x \cap G_y)\,] = \\ = [\,G : G_y\,] \cdot [\,G_y : (G_x \cap G_y)\,]$$

и переходя к логарифмам, получаем требуемое.

Пояснение к доказательству

Пусть $F \leqslant H \leqslant G$.
Тогда по теореме Лагранжа

$$|G| = |H| \cdot [G : H],$$
$$|H| = |F| \cdot [H : F],$$
$$|G| = |F| \cdot [G : F]$$

и

$$[G : F] = \frac{|G|}{|F|} = \frac{|G| \cdot [H : F]}{|H|} = [G : H] \cdot [H : F].$$

Взаимная информация слова

Определение

Взаимной 0*-информацией слов* x *и* y, или *информацией, содержащийся в слове* x *относительно слова* y назовём величину

$$I_0(x:y) = I_0(x) + I_0(y) - I_0(x \times y) = \\ = I_0(x) - I_0(x/y) = I_0(y) - I_0(y/x).$$

При $I_0(x:y) = 0$ слова x и y называют *независимыми*.

Ясно, что $I_0(x:y) = I_0(y:x)$ и если слова независимы, то $I_0(x) = I_0(x/y)$ и $I_0(y) = I_0(y/x)$.

Взаимная информация слова в нашем примере

1. $\underline{n = 3,\ x = aab,\ y = 111}$:

$$I_0(x) = \log 3,\ I_0(y) = 0,\ I_0(x/y) = \log 3,\ I_0(y/x) = 0,$$

$$I_0(x:y) = I_0(x) - I_0(x/y) = \log 3 - \log 3 = 0,$$
$$I_0(y:x) = I_0(y) - I_0(y/x) = 0 - 0 = 0.$$

2. $\underline{n = 5,\ x = aabab,\ y = 11000}$:

$$I_0(x) = I_0(y) = \log 10,\ \ I_0(x/y) = I_0(y/x) = \log 3.$$

$$I_0(x:y) = I_0(y:x) =$$
$$= I_0(x) - I_0(x/y) = \log 10 - \log 3 = \log \frac{10}{3}.$$

Условная информация слова: вычисление

Пусть имеются

- алфавиты X и Y из q и k букв соответственно;
- слова $x \in X^n$ и $y \in Y^n$ (длины n) над ними;
- композиции слов — $(m_1, \ldots, m_q)$ и $(n_1, \ldots, n_k)$ соответственно.

Составим $q \times k$ таблицу $M = \| m_{ij} \|$, где m_{ij} — число замен i-го символа слова x на j-й символ слова y, $i = \overline{1, q}$, $j = \overline{1, k}$.

В таблице M суммы элементов строкам будут равняться $m_1, \ldots, m_q$, а по столбцам — $n_1, \ldots, n_k$.

Тогда, как можно показать, $I_0(y/x) =$

$$= \log \frac{m_1!}{m_{11}! \ldots m_{1k}!} + \log \frac{m_2!}{m_{21}! \ldots m_{2k}!} + \ldots + \log \frac{m_q!}{m_{q1}! \ldots m_{qk}!}.$$

Условная 0-информация слова: пример

Пусть
$$x = (a\,b\,a\,a\,c\,b\,c\,a) \in \{a, b, c\}^8,$$
$$y = (1\,2\,2\,1\,3\,2\,1\,3) \in \{1, 2, 3\}^8.$$

Композиции слов: x — $(4, 2, 2)$, y — $(3, 3, 2)$.

Таблица 3×3 замен букв — есть

$$M = \begin{array}{c|ccc|c} & 1 & 2 & 3 & \\ \hline a & 2 & 1 & 1 & 4 \\ b & & 2 & & 2 \\ c & 1 & & 1 & 2 \\ \hline & 3 & 3 & 2 & 8 \end{array}$$

Тогда

$$I_0(y/x) = \log \frac{4!}{2!\,1!\,1!} + \log \frac{2!}{2!} + \log \frac{2!}{1!\,1!} =$$
$$= \log 12 + 0 + \log 2 = \log 24.$$

Условная 0-информация слова: пример...

Производя те же вычисления по столбцам, получим

$$I_0(x/y) = \log\frac{3!}{2!\,1!} + \log\frac{3!}{2!\,1!} + \log\frac{2!}{1!\,1!} =$$
$$= 2\log 3 + \log 2 = \log 18 \approx 4{,}170.$$

Для безусловной информации I_0 получим значения

$$I_0(x) = \log\frac{8!}{4!\,2!\,2!} = \log 420, \quad I_0(y) = \log\frac{8!}{3!\,3!\,2!} = \log 560.$$

Взаимная информация:

$$I_0(x:y) = I_0(x) - I_0(x/y) = \log\frac{420}{18} = \log\frac{70}{3} =$$
$$= I_0(y) - I_0(y/x) = \log\frac{560}{24} = \log\frac{70}{3} \approx \log 23,33 \approx 4{,}54.$$

Неравенства для количества 0-информации

Утверждение

$$1)\ I_0(y/x) \leqslant I_0(y), \qquad 2)\ I_0(x \times y) \leqslant I_0(x) + I_0(y).$$

Напоминание. Для слова $x \in X^n$, $|X| = q$ с композицией $(m_1, \ldots, m_q)$ и действия S_n на X^n, порождающего классы эквивалентности $\sim_{S_n}$: $G_x = \mathrm{Stab}\,(x) \leqslant S_n$,
$|\,\mathrm{Orb}\,(x)| = |G|/|G_x| = [G : G_x]$,
$I_0(x) = \log|\mathrm{Orb}\,(x)| = \log \frac{n!}{m_1! \cdot \ldots \cdot m_q!}$.

Доказательство

Первое неравенство очевидно: $\mathrm{Orb}\,_{G_x}(y)$ *элемента* y *под действием подгруппы* $G_x = \mathrm{Stab}\,(x) \leqslant S_n$ *есть подорбита* $\mathrm{Orb}\,(y)$ *безусловной под действием группы* S_n.

Второе неравенство следствие первого и равенства для $I_0(x \times y)$.

Разделы

Отождествление букв слова

Если на алфавите X задана эквивалентность $\sim$, то буквы из одного смежного класса отождествляются.

Утверждение

В результате отождествления некоторых букв алфавита информация слова уменьшается: если x и x' — исходное и новое слова, то $I_0(x') = I_0(x) - I_0(x/x')$.

Пример. Пусть в слове $x = (a\,b\,a\,a\,c\,b\,c\,a)$ буквы b и c отождествляются, тогда $x' = (a\,b\,a\,a\,b\,b\,b\,a)$, а при вычислении $I_0(x/x')$ можно ограничиться таблицей замены

$$M = \begin{array}{c|cc|c} & b & c & \\ \hline b & 2 & 2 & 4 \end{array}$$

$I_0(x/x') = \log \frac{4!}{2!\,2!} = \log 6 \approx 2{,}58,$

$I_0(x') = I_0(x) - I_0(x/x') = \log \frac{420}{6} = \log 70 \quad (I_0(x) = \log 420).$

Прямой подсчёт: $I_0(x') = \log \frac{8!}{4!4!} = \log \frac{5 \cdot 6 \cdot 7 \cdot 8}{2 \cdot 3 \cdot 4} = \log 70.$

Оптимальное квантование данных

Пусть в результате наблюдений получена последовательность данных — вектор x, который требуется закодировать бинарным словом.

Естественно требовать, чтобы новое слово $x' \in \{0, 1\}$ содержало как можно больше информации об исходном слове, т.е. чтобы значение $I_0(x')$ было максимальным.
Это достигается при равенстве количеств 0 и 1 в x'.

Таким образом, приходим к *квантованию по принципу вариационного ряда*: упорядочиваем значения x по возрастанию и кодируем 0 первую половину отсортированных данных, а 1 — вторую.

Оптимальное квантование данных: пример...

Пусть в результате наблюдений получена последовательность данных $x = (0{,}16,\ 0{,}1,\ 0{,}7,\ 0{,}18,\ 1{,}15)$

и требуется закодировать их бинарным словом.

Составляем вариационный ряд, и кодируем в нём первые по порядку три значения нулём, а последние — единицей:

x	0,16	0,1	0,7	0,1	0,18	1,5
порядок	3	1	5	2	4	6
x'	0	0	1	0	1	1

Название подхода в общем случае — *квантильное кодирование*. Например, разбивая вариационный ряд на 4 приблизительно равных интервала, кодируем данные двухбитными словами

$$00,\ 01,\ 10,\ 11.$$

Разделы

Вычисление информативности признака

Прецедентная информация задач распознавания образов может быть описана *матрицей информаций*, строки которой соответствуют объектам, столбцы — признакам, а дополнительный столбец содержит символы классов, которым принадлежат объекты.

Будем рассматривать задачу распознавания с качественными признаками, когда каждый признак принимает значения из множества $X = \{1, 2, \ldots, m\}$, множеством символов классов $Y = \{1, 2, \ldots, k\}$ и n прецедентами.

Информативность $I_0(x : y)$ признака $x \in X^n$ по отношению к *информационному вектору* символов классов $y \in Y^n$ объектов равна количеству информации, содержащимся в x относительно y:

$$I_0(x : y) = I_0(x) - I_0(x/y) = I_0(y) - I_0(y/x).$$

Вычисление информативности признака: пример 1

Пусть $x = (2,2,2,2,2,1,2,2,2)$ — значение признака и
$y = (1,1,1,2,2,2,3,3,3)$ — классификация у $n = 9$
прецедентов. Тогда

$$I_0(x) = \log\frac{9!}{8!\,1!} = \log 9 \text{ и } I_0(y) = \log\frac{9!}{3!\,3!\,3!} = \log 1680,$$

и таблица замен символов —

$$M = \begin{array}{c|ccc|c} & 1 & 2 & 3 & \\ \hline 1 & & 1 & & 1 \\ 2 & 3 & 2 & 3 & 8 \\ \hline & 3 & 3 & 3 & 9 \end{array}$$

$$I_0(y/x) = \log\frac{1!}{1!} + \log\frac{8!}{3!\,2!\,3!} = \log 560,$$

$$I_0(x/y) = 2\log\frac{3!}{3!} + \log\frac{3!}{1!\,2!} = \log 3,$$

$$I_0(x:y) = I_0(x) - I_0(x/y) = \log 9 - \log 3 = \log 3 \approx 1{,}585,$$

$$I_0(y:x) = I_0(y) - I_0(y/x) = \log 1680 - \log 560 = \log\frac{1680}{560} = \log 3.$$

Чистые классификаторы и чистые шумы

Определение

Признак x относительно информационного вектора символов классов y назовём

- *чистым классификатором*, если $I_0(x:y) = I_0(y)$ (т.е. $I_0(y/x) = 0$);
- *чистым шумом*, если $I_0(x:y) = 0$.

Ясно, что

- по значениям чистого классификатора информационный вектор восстанавливается полностью: в этом случае существует биекция $x \leftrightarrow y$;
- чистый шум не даёт никакой информации об информационном векторе (x и y независимы).

Нижний порог информативности признака

Для того, чтобы исключить «шумящие» признаки из дальнейшего рассмотрения, введём *нижний порог* α информативности признака и отбросим все признаки объектов с информативностью, меньшей α.

Нижний порог α удобно выбирать в процентах от максимально возможной информативности $I(y)$.

Пример ①. Для информационного вектора $y = (1,1,1,2,2,2,3,3,3)$ символов классов имеем: $y \in Y^9$, $Y = \{1,2,3\}$ и

$$I(y) = \log|Y^n| = \log 3^9 = \log 19683 \approx 14{,}26.$$

Например, если выбрать $\alpha = 24\%$, то все признаки x с информативностью $I_0(x:y) < 3{,}42 = 14{,}26 \cdot 0{,}24$ исключаются из дальнейшего рассмотрения.

Кластеры признаков: информационный подход

Утверждение

Функция

$$\rho(x^i, x^j) = \frac{1}{2}\left(I_0(x^i/x^j) + I_0(x^j/x^i) \right)$$

является метрикой на множестве признаков (= слов из X^n).

Ясно, что значение ρ между зависимыми (повторяющими друг друга признаками) близко к 0.

Назовём

- ρ — *информационной метрикой* на пространстве признаков;
- *кластером признаков* — подмножество близких друг к другу признаков (относительно метрики ρ).

Кластеры признаков: пример 1

Продолжение Примера ①.

Пусть информация о прецедентах в задаче распознавания задана матрицей с качественными признаками

	x^1	x^2	x^3	x^4	x^5	x^6	x^7	x^8	y
1	2	2	2	1	2	2	1	1	1
2	2	2	2	2	2	2	1	1	1
3	2	2	2	1	1	2	1	1	1
4	2	1	1	2	1	1	1	2	2
5	2	1	1	2	1	1	2	1	2
6	1	1	1	2	1	1	1	2	2
7	2	1	1	1	2	2	1	2	3
8	2	1	1	1	2	2	2	2	3
9	2	2	1	1	2	2	1	2	3

(значение $I(y) \approx 14{,}26$ вычислено ранее).

Кластеры признаков: пример 1 ...

1. Подсчитаем информативность $I_0(x^i : y)$, $i = \overline{1,8}$ признаков по отношению к информационному вектору y:

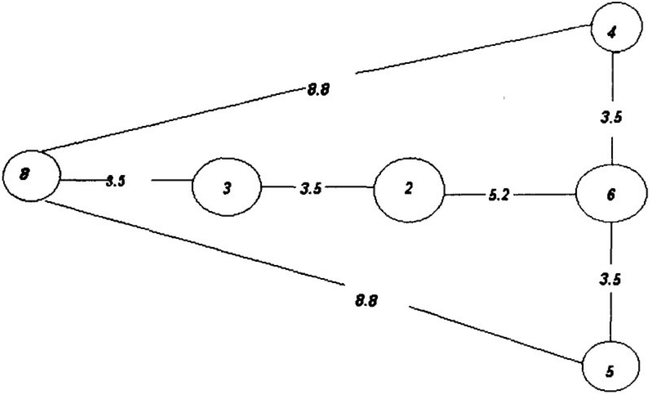
x^1	x^2	x^3	x^4	x^5	x^6	x^7	x^8
1.77	6.16	8.26	6.16	6.16	8.26	1.37	6.16

Приняв порог $\alpha = 24\%$, отбрасываем признаки x^1 и x^7 как малоинформативные (с $I_0(x : y) < 3.24$).

2. Составим таблицу расстояний оставшихся признаков:

ρ	x^2	x^3	x^4	x^5	x^6	x^8
x^2	0	3.5	8.4	8.4	5.2	6.8
x^3		0	8.4	8.4	6.0	3.5
x^4			0	6.8	3.5	8.8
x^5				0	3.5	8.8
x^6					0	8.4
x^8						0

Кластеры признаков: пример 1...



Анализ таблицы расстояний позволяет сделать вывод, что исследуемые признаки можно представить разбивающимися на два кластера: $\{x^2, x^3, x^8\}$ и $\{x^4, x^5, x^6\}$.

Кластеры признаков: пример 2

Пример ②. Прогнозирование запаса руды для месторождения.
Значения признаков бинарных признаков:

y — запас руды (1 — больше 1 млн тонн, 0 — меньше),
x^1 — приуроченность к горизонтам слюдистых сланцев,
x^2 — приуроченность к горизонтам амфиболитов,
x^3 — близость к контакту со слюдистыми сланцами,
x^4 — близость к контакту с амфиболитами,
x^5 — присутствие тел амфиболитов,
x^6 — обилие даек,
x^7 — пиритизация,
x^8 — пропилитизация,
x^9 — ожелезнение,
x^{10} — аргиллитизация,
x^{11} — наличие вторичных ореолов свинца и цинка,
x^{12} — развитие складок.

Кластеры признаков: пример 2 ...

	x^1	x^2	x^3	x^4	x^5	x^6	x^7	x^8	x^9	x^{10}	x^{11}	x^{12}	y
1	0	1	0	0	0	1	1	1	0	0	1	1	1
2	0	1	1	1	0	0	1	0	1	0	1	1	1
3	0	1	0	1	1	1	0	1	0	0	0	0	1
4	0	0	1	0	1	0	1	0	1	1	1	1	1
5	1	1	0	1	0	1	1	0	1	0	1	0	1
6	1	0	1	0	1	1	0	1	0	1	0	1	0
7	1	0	1	0	1	1	1	0	0	1	0	0	0
8	1	0	0	1	1	1	0	1	0	0	0	0	0
9	1	0	1	1	0	0	0	0	1	0	0	1	0
10	0	1	1	0	1	1	0	1	0	1	0	1	0
11	0	0	1	1	1	1	0	0	0	0	0	1	?
12	0	0	0	1	0	1	0	0	1	0	1	1	?

Кластеры признаков: пример 2 ...

1. Вычислим информативность признаков ($I(y) = \log 2^{10} = 10$):

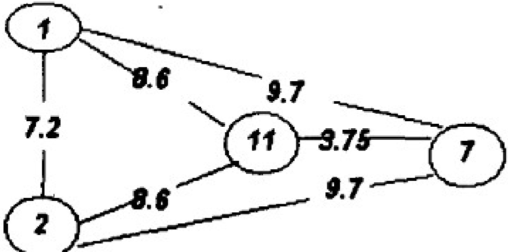
x^1	x^2	x^3	x^4	x^5	x^6	x^7	x^8	x^9	x^{10}	x^{11}	x^{12}
2,8	2,8	1,24	0,3	1,24	0,3	2,8	0,3	1,24	1,24	6,1	0

Если выбрать $\alpha = 25\%$, то останутся признаки x^1, x^2, x^7 и x^{11} (с информативностями $\geqslant 2{,}5$ относительно y).

2. Составим таблицу расстояний этих признаков:

ρ	x^1	x^2	x^7	x^{11}
x^1	0	7,2	9,7	8,6
x^2		0	9,7	8,6
x^7			0	3,75
x^{11}				0

Кластеры признаков: пример 2 ...



Анализ таблицы позволяет провести следующую кластеризацию признаков: $\{\, x^7, x^{11} \,\}$, $\{\, x^1 \,\}$, $\{\, x^2 \,\}$.

Сложные признаки

Если x^i и x^j — признаки, то признак $x^{i,j} = x^i \times x^j$ назовём *сложным*.

Утверждение

Информативность сложного признака $x^i \times x^j$ по отношению к информационному вектору символов классов y не меньше суммы их информативностей, точнее

$$I_0(y : (x^i \times x^j)) \geqslant I_0(y : x^i) + I_0(y : x^j) - I_0(x^i : x^j).$$

Поскольку значение $I_0(x^i : x^j)$ монотонно возрастает (от 0) при увеличении зависимости признаков x^i и x^j (суммарной длины фрагментов, получающихся перекодировкой), справедливо следующее правило:

при образовании сложного признака следует объединять признаки, лежащие в разных кластерах

Формирование сложных признаков

Введём второй параметр настройки β, равный минимально допустимой информативности сложного признака.
Величину β выбирают не менее 60% от $I_0(y)$.

В Примере ① было выделено два кластера признаков: $\{x^2, x^3, x^8\}$ и $\{x^4, x^5, x^6\}$, при $\beta = 80\%$ получаем порог $0{,}8 \cdot 14{,}26 = 11{,}4$.
Вычисляя значение $I_0(y : (x^i \times x^j))$ для признаков из разных классов, найдём, что информативность сложных признаков

$$\begin{aligned} x^{3,6} &\ \text{— достаточна,} \\ x^{2,4} &\ \text{— недостаточна,} \\ x^{2,4,8} &\ \text{— достаточна.} \end{aligned}$$

Формирование сложных признаков: пример 2

В Примере ② положим $\beta = 60\%$, что установит порог информативности $0{,}6 \cdot 10 = 6$ для формируемых сложных признаков.

Образуем сложный признак $x^{1,11} = x^1 \times x^{11} = x$.
Его взаимную относительно информационного вектора y информативность $I_0(x : y)$ определим по таблице переходов значений x в значения y, которая, в свою очередь, заполняется строится по соответствующим столбцам матрицы информаций.

Формирование сложных признаков: пример 2...

x^1	x^{11}	y
0	1	1
0	1	1
0	0	1
0	1	1
1	1	1
1	0	0
1	0	0
1	0	0
1	0	0
0	0	0

$$I_0(x:y) = I_0(x) - I_0(x/y),$$

	0	1	
01		3	3
00	1	1	2
11		1	1
10	4		4
	5	5	10

$$I_0(x) = \log \frac{10!}{3!\,2!1!\,4!} = \log 12600,$$

$$I_0(x/y) = \log \frac{5!}{1!\,4!} + \log \frac{5!}{3!\,1!\,1!} = \log 100,$$

$$I_0(x:y) = \log \frac{12600}{100} = \log 126 \approx 6{,}98 > 6,$$

т.е. сложный признак $x = x^1 \times x^{11}$ удовлетворяет условию информативности.

Формирование сложных признаков: пример 2...

Аналогично находим, что информативность, больше 6 имеют сложные признаки $x^{2,11}$, $x^{1,2,7}$ и $x^{7,11}$ и осуществляем переход в новое признаковое пространство полученных сложных признаков.

Значения $f(x)$ сложного признака x данного объекта определяют в зависимости от третьего параметра настройки γ, принимающего целочисленные значения и выражающего «степень значимости» признака.

Обозначим через n_1 и n_0 число единичных и, соответственно, нулевых значений информационного вектора, соответствующих данному значению полученного сложного признака x.
Полагаем

$$f(x) = \begin{cases} 1, & \text{если } n_1 - n_0 \leqslant \gamma, \\ 0, & \text{если } n_0 - n_1 \leqslant \gamma, \\ -, & \text{иначе (отказ от распознавания).} \end{cases}$$

Формирование сложных признаков: пример 2...

В рассматриваемом примере примем $\gamma = 2$.

Для признака $x = x^{1,11}$ имеем (дублируем таблицу):

x	0	1		$f(x)$
01		3	3	1
00	1	1	2	–
11		1	1	–
10	4		4	0

Тогда, например, для 1, 11 и 12-го объектов получим

$f_1(x) = 1$ (справочно),

$f_{11}(x) = -$, $f_{12}(x) = 1$.

Формирование сложных признаков: пример 2...

Для признака $x = x^{11,2}$:

x^{11}	x^2	y
1	1	1
1	1	1
0	1	1
1	0	1
1	1	1
0	0	0
0	0	0
0	0	0
0	0	0
0	1	0

x	0	1		$f(x)$
11		3	3	1
01	1	1	2	–
10		1	1	–
00	4		4	0
	5	5	10	

$I_0(y : (x^{11} \times x^2)) = 8$ бит

Формирование сложных признаков: пример 2...

Для признака $x = x^{1,2,7}$:

x^1	x^2	x^7	y
0	1	1	1
0	1	1	1
0	1	0	1
0	0	1	1
1	1	1	1
1	0	0	0
1	0	1	0
1	0	0	0
1	0	0	0
0	1	0	0

x	0	1		$f(x)$
011		2	2	1
010	1	1	2	–
001	0	1	1	–
111	0	1	1	–
100	3	0	3	0
101	1	0	1	–
	5	5	10	

$I_0(y : (x^{11} \times x^2)) = 8$ бит

Формирование сложных признаков: пример 2...

Для признака $x = x^{11,7}$:

x^{11}	x^{7}	y
1	1	1
1	1	1
0	0	1
1	1	1
1	1	1
0	0	0
0	1	0
0	0	0
0	0	0
0	0	0

x	0	1		$f(x)$
11	0	4	4	1
00	4	1	5	0
01	1	0	1	–
	5	5	10	

$$I_0(y : (x^{11} \times x^2)) = 6{,}4 \text{ бит}$$

Формирование сложных признаков: пример 2...

Классификацию будем осуществлять по близости (в метрике Хэмминга) к векторам классов и в новом признаковом пространстве имеем:

Номер объекта	$x^{1,11}$	$x^{2,11}$	$x^{1,2,7}$	$x^{7,11}$	y
11	–	0	–	0	0
12	1	1	–	–	1

Разделы

Литература I

- *Гоппа В.* Введение в алгебраическую теорию информации. — М.: Наука, 1995.
- *Верещагин Н. К., Щепин Е. В.* Информация, кодирование и предсказание. — М.: ФМОП, МЦНМО, 2012.