# Алгоритмы распознавания, основанные на оценке взаимосвязанности признаков

***Камилов М. М., Фазылов Ш. Х., Мирзаев Н. М.***
kamilov@yandex.ru, shavkat-faz@mail.ru, mnm2005@rambler.ru
Ташкент, Институт математики и информационных технологий АН Руз

В данной работе рассматривается вопрос построения алгоритмов распознавания на основе анализа взаимосвязанности признаков, которые являются логическим продолжением работ Ю. И. Журавлева и его учеников [1–4]. Корректность, и другие не приведенные здесь определения понятий, а также обозначения, можно найти в [1, 2]. Задание этих алгоритмов включает следующие основные этапы.

**1. Задание меры парных связей между признаками.** Пусть $S$ — совокупность допустимых объектов. В пространстве признаков $X = (x_1, \ldots, x_n)$ любому объекту $S \in \{S\}$ соответствует вектор $\bar{a} = (a_1, \ldots, a_n)$. Каждый признак $x_i$ имеет своё множество допустимых значений $D_i,\ i = 1, \ldots, n$. Введём функцию $\mu(x_i, x_j)$, характеризующую силу парной связи между признаками $x_i$ и $x_j$, и удовлетворяющую условиям:

1) $\mu(x_i, x_j) \geqslant 0$;
2) $\mu(x_i, x_j) = \mu(x_j, x_i)$;
3) $\mu(x_i, x_i) > \mu(x_i, x_j), \quad i \neq j$.

В зависимости от выбора $\mu(x_i, x_j)$ можно получить разнообразные алгоритмы определения взаимозависимости признаков.

**2. Определение «независимых» множеств сильносвязанных признаков.** Пусть $A_k,\ k = 1, \ldots, k_0$ — множества сильносвязанных признаков. Меру близости $L(A_p, A_q)$ можно задать различными способами, например:

$$L(A_p, A_q) = \frac{1}{N_p N_q} \sum_{x_i \in A_p} \sum_{x_j \in A_q} \mu(x_i, x_j),$$

где $N_p, N_q$ — число признаков, входящих, соответственно, в множества $A_p$, $A_q$. В зависимости от способа задания меры близости $L(A_p, A_q)$ между $A_p$ и $A_q$ можно получить разнообразные алгоритмы выделения независимых групп сильносвязанных признаков.

**3. Определение моделей функциональной зависимости в каждой группе признаков для каждого класса $K_j,\ j = 1, \ldots, l$.** Пусть $x_i$ — произвольный признак из группы $A_q$, и $x_{i_0} = \arg\max\limits_{x_i} \sum\limits_{x_i \in A_q} \mu(x_i, x_j)$. Тогда модели функциональной зависимости в $A_q$ зададим в виде

$$x_{i_0} = F(\bar{c}, \bar{y}), \quad \bar{y} \in A_q \setminus \{x_{i_0}\},$$

где $\bar{c}$ — вектор неизвестных параметров, $F$ — функция из некоторого заданного класса $\{F\}$.

Вычисленные значения вектора неизвестных параметров $\bar{c}$ определяют модель функциональной зависимости. В зависимости от задания параметрического вида $F(\bar{c}, \bar{x})$ и метода определения $\bar{c}$ получим разнообразные алгоритмы распознавания.

**4. Определение элементарных пороговых правил принятия решений.** Сформулируем элементарные пороговые правила принятия решений, основанные на задании порогов. Они характеризуют допустимые отклонения в рассматриваемой модели функциональной зависимости. Для простоты рассмотрим два класса объектов. Обозначим через $K_j$ и $\complement K_j$ множества объектов, данных для обучения.

Определим множество $A_q$ всех функциональных моделей во множестве признаков $A_q$. Обозначим через $\delta_i$ все элементарные пороговые правила принятия решений в $A_q$:

$$\delta_i(K_j, S) = \begin{cases} 1, & \text{если } \left|x_{i0} - F(\bar{c}, x_i)\right| < \varepsilon_i; \\ 0, & \text{в противном случае.} \end{cases}$$

где $\varepsilon_i$ — заданный порог близости в рамках модели $x_{i_0} = F(\bar{c}, x_i)$.

**5. Задание функции близости $B_q(K_j, S)$ между $K_j$ и $S$ по множеству признаков $A_q$.** Функцию близости $B_q(K_j, S)$ можно задавать различними способами, например:

$$B_q(K_j, S) = \sum_{i=1}^{N_q} \rho_i \delta_i(K_j, S),$$

где $N_q$ — число используемых элементарных проговых правил. $\rho_i$ — параметр алгоритма.

**6. Оценка принадлежности объекта к классу.** Оценка принадлежности объекта к классу $K_j$, $j = 1, \ldots, l$, вычисляется оператором $B(S) = (\eta_1(S), \ldots, \eta_l(S))$, где

$$\eta_j(S) = \sum_{u=1}^{k_0} \gamma_u B_u(K_j, S),$$

$\gamma_u$ — параметр алгоритма, $u = 1, \ldots, k_0$; $k_0$ — число независимых множеств.

**7. Решающее правило.** Решение принимается поэлементно [1], т. е.

$$\beta_i = \begin{cases} 0, & \text{если } \eta_j(S_i) < c_1; \\ 1, & \text{если } \eta_j(S_i) > c_2; \\ \Delta, & \text{если} c_1 \leqslant \eta_j(S_i) \leqslant c_2. \end{cases}$$

Нами определены параметрические модели распознающих алгоритмов, основанные на оценке взаимосвязанности признаков. Всякий алгоритм $A$ из этой модели полностью определяется заданием набора параметров $\pi = \langle \bar{c}, \{\varepsilon_i\}, N_q, \{\rho_i\}, k_0, \{\gamma_u\} \rangle$. Множество всех алгоритмов распознавания из рассмотренной модели обозначим через $A(\pi, S)$. Очевидно, что существенным недостатком таких эвристических алгоритмов распознавания, проверяемых лишь на некотором количестве практических задач, является то, что алгоритмы, оптимальные для решения одной задачи из заданного класса, не всегда оптимальны (или приемлемы) для решения другой задачи того же класса. Поэтому возникает задача исследования корректности рассмотренных алгоритмов. В связи с этим доказана следующая

**Теорема 1.** *Пусть множество $\{Z\}$ удовлетворяет условиям:*
*1) объекты, изоморфные относительно $J_0$, отсутствуют в $\tilde{S}^q$;*
*2) оператор $B(S)$ ограничен;*
*3) $J_0 \cap \tilde{S}^q = \varnothing$.*

Тогда в рамках алгебраического замыкания расмотренных алгоритмов существует корректный алгоритм $A^*$ для задачи $Z \in \{Z\}$.

### Литература


[1] *Журавлев Ю. И.* Корректные алгебры над множествами некорректных (эвристических) алгоритмов // Кибернетика. — 1977. №4. — С.14–20.

[2] *Журавлев Ю. И., Исаев И. В.* Построение алгоритмов распознавания, корректных для заданной контрольной выборки // ЖВМиМФ. — Москва, 1979. — Т. 19. — №3. — С. 726–738.

[3] *Журавлев Ю. И., Камилов М. М., Туляганов Ш. Е.* Алгоритмы вычисления оценок и их применение. — Ташкент: Фан, 1974. — 119 с.

[4] *Фазылов Ш. Х., Мирзаев Н. М., Жуманазаров С. С.* О корректности алгоритмов распознавания, основанных на взаимосвязанности между признаками // «Проблемы информатики и энергетики». — Ташкент, 1997. — №1. — С. 19–24.