УДК 621.319:519.2.

ТАКСОНОМИЯ В БУЛЕВОМ ПРОСТРАНСТВЕ

В.И. Котюков

Данная работа является по существу продолжением работы [1].

Задача таксономии рассматривается для случая "двоичных шкал наименований" $\{X_i\}$, где $X_i=\{1,0\}$, $i=1,\ldots,p$.

Под <u>таксоном</u> будем понимать множество объектов, на котором выполняется некоторая закономерность F_j распределения значений переменных $\{X_i\}$.

Общую таксономическую решающую функцию F можно представить в виде некоторой суперпозиции функций F_j полученных таксонов:

$$F=\varphi(F_j(x)) \quad , \text{ где}$$

$$F_j(x)=\begin{cases} 1, & x\in S_j \\ 0, & x\bar{\in} S_j \end{cases} ; \qquad (1)$$

m - число таксонов S_j. $\quad j=1,\ldots m;$

Величина m может быть как известной заранее, так и неизвестной.

Пусть H - множество всех вершин n-мерного гиперкуба, где n - число переменных X_i, на которых в результате обучения оказалась определена F.

Заметим, что $n\leq p$.

К результату процесса обучения будет предъявлено такое <u>требование</u> :

если $F_i(x)=1$, то $F_j(x)=0$ и $\bigvee_{i=1}^{m} F_i(x)=1$, $\qquad (2)$

где $x \in H$ и $i \neq j$; $i, j = 1, \ldots, m$.

Как и ранее [I], будем исходить из предположения, что для "булевых шкал наименований" наиболее характерны закономерности, описываемые на языках булевой алгебры и метрики Хэмминга.

Будем полагать, что F_j есть "свойство" (смотри [I]) выборки LS_j.

Поскольку все исходные предпосылки здесь одинаковы с задачей распознавания, мы для наших целей можем использовать аппарат метода "обобщающихся закономерностей" ("условных корреляций"), развитый в [I].

Вернемся к задаче распознавания. В результате построения решающего правила F мы каждый класс (вид, таксон) описывали как совокупность некоторых подклассов (подвидов, подтаксонов) $\{a_\ell, f_\ell\}, \ell = 1, \ldots, k$, где f_ℓ – некоторые "свойства", а a_ℓ – соответствующие им условия.

В задаче таксономии мы можем считать всю общую выборку одним таксоном. Тогда наша задача будет заключаться в выделении подтаксонов $\{a_\ell, f_\ell\}$, которые мы в нашей задаче и будем называть просто таксонами.

Таким образом, в нашем случае таксон F_ℓ будет эквивалентен условию a_ℓ $(F_\ell = a_\ell)$, при котором найдено некоторое свойство f_ℓ. Свойство f_ℓ будет у нас как бы "описанием" таксона, а условие a_ℓ – его "именем", по которому мы его и будем далее определять.

Если мы имеем $F_\ell = a_\ell$, $\ell = 1, \ldots, m$, то нетрудно убедиться, что требование (2) будет выполнено, ибо условия $\{a_\ell\}$ отвечают "принципу дополнения" [I].

Свойства $\{f_\ell\}$ различных таксонов должны быть различны.

Процесс дробления пространства X при поиске закономерностей имеет древовидную структуру. Из всех возможных минимальных деревьев, отвечающих "принципу дополнения" (решающих функций F), предпочтение будет отдано тому, у которого максимальна доверительная вероятность его надежности:

$$\left(\gamma = \sum_{\ell=1}^{m} \bar{P}(f_\ell) \cdot \beta(f_\ell) \cdot \gamma(f_\ell)\right)_{max}.$$

Подробнее о вычислении величины γ сказано в [I].

Литература

I. Котюков В.И. Распознавание образов в булевом пространстве. (данный сборник).

Поступила в редакцию 21.I.1971 г.