ТАКСОНОМИЯ В БУЛЯМ ПРОСТРАНСТВЕ

В.И. Косыков

Данная работа является по существу продолжением работы [1].
Задача таксономии рассматривается для случая "двоичных" имен
наименований \( \{ x_i \} \), где \( x_i = \{ 0, 1 \} \), \( i = 1, \ldots, \rho \).
Под таксономией будем понимать множество объектов, на кото-
ром выполняется некоторая закономерность \( F_j \) распределения
значений переменных \( \{ x_i \} \).
Общую таксономическую решающую функцию \( F \) можно предста-
вить в виде некоторой суперпозиции функции \( F_j \) полученных так-
сонов:
\[
F = \varphi \left( F_j (x) \right) \quad \text{где}
\]
\[
F_j (x) = \begin{cases} 
1, & x \in S_j \\
0, & x \in \overline{S}_j 
\end{cases}
\]
(1)
\[
m \quad \text{число таксонов} \quad S_j \quad j = 1, \ldots, m ;
\]
Величина \( m \) может быть как известной заранее, так и
неизвестной.
Пусть \( \mathcal{H} \) множество всех вершин \( \rho \)-мерного гиперкуба,
где \( \rho \) - число переменных \( x_i \), на которых в результате обу-
чения оказалась определена \( F \).
Заметим, что \( \rho \leq \rho \).
К результату процесса обучения будет предъявлено такое транс-
формирование:
если \( F_j (x) = 1 \), то \( F_j (x) = 0 \) и \( \bigvee_{i=1}^{m} F_i (x) = 1 \),
(2)
где \( x \in H \) и \( i, j; i, j = 1, \ldots, m \).

Как и ранее [I], будем исходить из предположения, что для "булевых значений на множестве" наиболее характерны закономерности, описывающие на языке булевой алгебры а метрики Дэвиды.

Будем использовать, что \( F_0 \) есть "плотность" (смотря [I]) выбора \( L \).

Поскольку все рассмотренные ранее одинаковые задачи рациональности, мы для наших целей можем использовать аппарат метода "ошибок между закономерностями" ("условных корреляций"), развитый в [I].

Вернемся в задачу восстановления. В результате построения решаемого правила \( F_0 \) мы находим класс (вид, таксон) искомых как совокупность некоторых подклассов (подвидов, подтаксонов) \( \{ a_i, f_e \}, i = 1, \ldots, n \), где \( f_e \) — некоторые "свойства", а \( a_i \) — соответствующие им условия.

В задаче восстановления, мы можем считать этой общей выборкой одним таксоном. Тогда наша задача будет заключаться в выделении подтаксонов \( \{ a_i, f_e \} \), которые мы в нашей задаче и будем называть просто таксонами.

Таким образом, в нашем случае таксон \( F_0 \) будет эквивалентным условию \( a_i (F_0 - a_i) \), при котором найдено некоторое свойство \( f_e \). Свойство \( f_e \) будет у нас как бы "описанием" таксона, а условие \( a_i \) — его "именем", по которому мы его и будем далее определять.

Если мы имеем \( F_0 = a_i, f_e = \ldots, m \), то нетрудно убедиться, что требование (2) будет выполнено, ибо условие \( a_i \) отвечает "принципу дополнения" [I].

Свойства \( f_e \) различных таксонов должны быть различными.

Процесс дробления пространства \( X \) при помощи закономерностей имеет древовидную структуру. Из всех возможных минимальных деревьев, отвечающих "принципу дополнения" (решающих функций \( F_0 \)), предпочтение будет отдано тому, у которого максимум "малых" доверительной вероятности его надежности:

\[
(\gamma = \sum_{f_e} \beta(f_e) \beta(f_e) \gamma(f_e))_{\max}
\]

Подробнее о вычислении величины \( \gamma \) сказано в [I].

Литература

1. Котков Б.И. Распознавание образов в булевом пространстве. (Данный сборник).

Поступила в редакцию 21.11.1971 г.