

## Критерии оптимизации выбора безызбыточных диагностических тестов для принятия решений в интеллектуальных диагностических системах

Янковская А. Е.

yank@tsuab.ru

Томск, Томский архитектурно-строительный университет

Безызбыточные (тупиковые [1]) безусловные диагностические тесты (ББДТ) используются для отнесения исследуемого объекта к одному из распознаваемых образов [1]. От свойств используемых тестов существенно зависит качество тестового распознавания. Однако выбор «хороших» тестов не всегда приводит к хорошим решениям, поскольку общее количество признаков в выбранном множестве тестов может оказаться слишком большим, также как временные и стоимостные затраты или ущерб, наносимый в результате выявления значений признаков, например, при решении геоэкологических и медицинских задач. Кроме того, на качество тестов влияет способ вычисления весовых коэффициентов признаков, входящих в тесты, используемые для принятия итогового решения.

Проблема нахождения «лучших» ББДТ, позволяющих синтезировать на их основе более простые правила принятия решений [1, 2, 3], не потеряла своей актуальности. В работе [4] впервые была поставлена задача выбора множества ББДТ с заданными свойствами и предложен алгоритм её решения. Количество сформулированных автором критериев выбора оптимального подмножества ББДТ постепенно увеличивалось и достигло 5 в статье [5]. В настоящем докладе формулируется 6 критериев и постановка задачи выбора оптимального множества ББДТ. Обосновывается целесообразность введения нового критерия выбора.

### Основные понятия и определения

Воспользуемся определениями и обозначениями, введенными в работах [2, 5].

*Тестом* называется совокупность признаков, различающих любые пары объектов, принадлежащих разным образам. Тест называется *безызбыточным*, если при удалении любого признака тест перестает быть тестом. Признак называется *обязательным*, если он содержится во всех ББДТ, и *псевдообязательным*, если он не является обязательным и входит в множество используемых при принятии решений ББДТ.

Обозначим через  $N = \{N_1, \dots, N_n\}$  — множество тестов;  $Z = \{z_1, \dots, z_m\}$  — множество признаков;  $L_i$  — множество признаков, входящих в тест  $N_i$ ,  $i \in \{1, \dots, n\}$ ;  $\mathbf{T}$  — булева матрица тестов, строки которой сопоставлены тестам  $N_i$ , где  $N_i \in N$ ,  $i \in \{1, \dots, n\}$ , столбцы  $x_j$  сопоставлены признакам  $z_j \in Z$ ,  $j \in \{1, \dots, m\}$ ;  $n_0$  — число используе-

мых для принятия решений тестов;  $\mathbf{T}_0$  — подматрица матрицы  $\mathbf{T}$ ;  $N_0$  — множество тестов, соответствующих строкам матрицы  $\mathbf{T}_0$ ,  $N_0 \subseteq N$ .

Обозначим через  $w_j^r$  ( $w_j^g$ ) весовой коэффициент признака  $z_j \in Z$ ,  $j \in \{1, \dots, m\}$ , определяемый как разделяющая способность признака [2] (информационный вес по формуле  $w_j^g = \frac{k_j}{n_0}$ , где  $k_j$  — количество единичных значений в столбце  $x_j$  матрицы  $\mathbf{T}_0$ ).

Каждому тесту  $N_i$ ,  $i \in \{1, \dots, n\}$  соответствуют:

- 1) вес теста  $W_i^r = \sum_{j \in L_i} w_j^r$ ;
- 2) вес теста  $W_i^g = \sum_{j \in L_i} w_j^g$ ;
- 3) стоимость теста  $W_i^s = \sum_{j \in L_i} w_j^s$ , где  $w_j^s$  — коэффициент стоимости признака  $z_j \in Z$ ,  $j \in \{1, \dots, m\}$ ;
- 4) риск  $W_i'' = \sum_{j \in L_i} w_j''$ , где  $w_j''$  — ущерб (риск), наносимый в результате выявления (измерения) значения  $j$ -го признака.

#### Постановка задачи. Критерии оптимальности

Дано множество тестов  $N$ , представленное матрицей  $\mathbf{T}$ ; множество признаков  $Z$ , каждый из которых содержится хотя бы в одном тесте из  $N$ ; а также: весовые коэффициенты  $w_j^r$  и  $w_j^g$ , коэффициенты стоимости признаков  $w_j^s$  и величина ущерба (риска)  $w_j''$ ,  $j \in \{1, \dots, m\}$  от выявления значения  $j$ -го признака; веса  $W_j^r$  и  $W_j^g$ , стоимости  $W_j^s$ , ущерб (риск)  $W_j''$ ,  $i \in \{1, \dots, n\}$  тестов.

Необходимо выделить из матрицы  $\mathbf{T}$  такую подматрицу  $\mathbf{T}_0$ , содержащую  $n_0$  строк, чтобы соответствующее ей выделенное множество тестов  $N_0$  обеспечивало выполнение следующих критериев оптимальности:

- 1) во множестве тестов  $N_0$  должно содержаться максимальное число псевдообязательных признаков;
- 2) множество тестов  $N_0$  должно содержать минимальное общее число признаков;
- 3) множество тестов  $N_0$  должно иметь максимальный суммарный вес  $W_0^r = \sum_{i \in N_0} W_i^r$ ;
- 4) множество тестов  $N_0$  должно содержать максимальный суммарный вес  $W_0^g = \sum_{i \in N_0} W_i^g$ ;
- 5) множество тестов  $N_0$  должно иметь наименьшую суммарную стоимость;
- 6) множество тестов  $N_0$  должно обеспечивать наименьший ущерб (риск).

Заметим, что критерий 4) аналогичен критерию по выбору признаков с максимальным информационным весом [1], вычисляемым только не на всём множестве ББДТ, а на множестве  $N^0$ , используемом для принятия итогового решения.

**Утверждение 1.** Информационный вес  $W_0^g$  множества тестов  $N^0$ , умноженный на  $n_0$ , равен количеству  $K_{n_0}$  единичных значений в матрице  $T_0$ .

**Доказательство.** Поскольку  $W_0^g = \sum_{j=1}^{n_0} w_j^g = \sum_{j=1}^{n_0} \frac{k_j}{n_0} = \frac{K_{n_0}}{n_0}$ , следовательно,  $W_0^g n_0 = K_{n_0}$ , что и требовалось доказать.

Критерий 4) можно переформулировать следующим образом: множество тестов  $N^0$  должно включать максимальное количество признаков.

Дальнейшее расширение списка критериев выбора связано с учетом зависимости входящих в тесты признаков.

Модификация приведенных в [5] трех алгоритмов (логико-комбинаторного, на основе метода анализа иерархий, генетического) выбора оптимального множества тестов с учетом критерия 4) не представляет затруднений. Модификация алгоритмов будет воплощена в интеллектуальном инструментальном средстве ИМСЛОГ [6].

Работа выполнена при поддержке РФФИ, проект № 07-01-00452а и РГНФ, проект № 06-06-12603в.

### Литература

- [1] Журавлев Ю. И., Гуревич И. Б. Распознавание образов и анализ изображений // Искусственный интеллект: В 3-х кн. Кн.2. Модели и методы: Справ./ Под ред. Д.А.Поспелова. — М.: Радио и связь, 1990. — С. 149–191.
- [2] Янковская А. Е. Логические тесты и средства когнитивной графики в интеллектуальной системе // Новые информационные технологии в исследовании дискретных структур: Докл. 3-ей Всерос. конф. с межд. участ., Томск: Изд-во СО РАН, 2000. — С. 163–168.
- [3] Naidenova R. A., Plaksin M. V., Shagalov V. L. Inductive inferring all good classification test // Знание-Диалог-Решение. Сб. науч. тр. Межд. конф. Том 1. Ялта, 1995. — С. 79–84.
- [4] Янковская А. Е. Построение логических тестов с заданными свойствами и логико-комбинаторное распознавание на них // Интеллектуализация обработки информации. Тезисы докл. — Симферополь, 2002. — С. 100–102.
- [5] Колесникова С. И., Можейко В. И., Цой Ю. Р., Янковская А. Е. Алгоритмы выбора оптимального множества безыбыточных диагностических тестов в интеллектуальных системах поддержки принятия решений // Труды первой межд. конф. САИТ-2005. — Т. 1. — М.: КомКнига, 2005. — С. 256–262.

- 
- [6] *Yankovskaya A. E., Gedike A. I., Ametov R. V., Bleikher A. M.* IMSLOG-2002 Software Tool for Supporting Information Technologies of Test Pattern Recognition // Pattern Recognition and Image Analysis. — 2003. — Vol. 13, №4. — P. 650–657.