

Теория тестового распознавания*

В. Б. Кудрявцев, А. Е. Андреев

Описывается логический подход к распознаванию образов, его основным понятием выступает тест. Анализ совокупности тестов позволяет строить функционалы, характеризующие образ и процедуры вычисления их значений. Указываются качественные и метрические свойства тестов, функционалов и процедур распознавания. Приводятся результаты решения конкретных задач.

Введение

Многие задачи распознавания могут быть описаны с помощью следующей схемы. Имеется некоторый объект A , который может находиться в каких-то состояниях q_i , $i \in \{1, 2, \dots, m\}$. Последние характеризуются параметрами-признаками x_j , $j \in \{1, 2, \dots, n\}$. Известно, что некоторые наборы значений этих признаков $\alpha = (a_1, a_2, \dots, a_n)$ описывают состояние q_i . Эти наборы как строки составляют матрицу T_i . Совокупность T как матрица, образованная подматрицами T_i , описывает все состояния объекта A . Для того, чтобы понять, какое состояние описывает конкретный набор значений признаков, нужно проверить, в какую подматрицу T_i он входит. Если он не содержится в них, то набор не соответствует состояниям. Если входит в подматрицу T_i , то в предположении, что подматрицы не пересекаются, этот набор описывает состояние q_i .

В реальной ситуации и признаки, и сами подматрицы T_i , как правило, описываются не полностью, а потому решение вопроса о том,

*Работа поддерживалась грантом РФФИ № 06-01-00240.

какое состояние описывает данный набор, становится нетривиальным. Этот вопрос и называют задачей распознавания.

Тем самым задачу распознавания можно сформулировать так: даны матрица T' и ее подматрицы T'_i , требуется указать оператор φ , который по этим подматрицам и заданному набору признаков вычисляет состояние объекта A , представленное предъявленным набором. Ясно, что нахождение φ в общем случае требует различных допущений относительно свойств A .

К числу исторически первых таких допущений относятся алгебраические, геометрические и вероятностные свойства A . Каждое из них породило соответствующее направление в теории распознавания. В них накоплен большой опыт, включающий как перечень решенных и потенциально решаемых задач, так и методов их решения.

Особый класс составляют задачи, в которых характеристика объекта A является опосредованной или абстрактной, без подходящей интерпретации.

Типичным примером такого объекта является техническое устройство, характеризуемое некоторыми своими параметрами-признаками. Это устройство может иметь неисправности — состояния, каждое из которых описывается соответствующей подматрицей T_i . Тогда, зная T , по текущему набору признаков $\alpha = (a_1, a_2, \dots, a_n)$, можно, как отмечалось, решить задачу распознавания состояния q объекта A .

С. В. Яблонским и И. А. Чегис [1] было замечено, что вообще говоря, такое распознавание можно осуществлять, не используя всю матрицу T , а только ее часть. Ими было введено понятие теста для T следующим образом.

Пусть σ — набор признаков, $T_{i,\sigma}$ — часть таблицы T_i , образованная столбцами, соответствующими σ , а T_σ — все $T_{i,\sigma}$. Набор σ образует тест для T , если $T_{i,\sigma}$ и $T_{i',\sigma}$ не имеют общих строк при $i \neq i'$. Таким образом, тест σ уже сам может решать задачу распознавания после анализа множества T_σ и набора α_σ . Тест выступает в роли эксперта, принимающего решение по части набора α и по T .

Авторы предложили логическое решение задачи описания множества $\mathfrak{Z}(T)$ всех тестов для T . Этот подход нашел применение в технической диагностике.

Позже он был распространен и на другие объекты. В качестве A были рассмотрены рудные образования в предположении, что T доступно лишь фрагментарно в виде T' , а множество тестов для T и T' , видимо, «мало» отличаются друг от друга.

Ю. И. Журавлевым, Ф. П. Кренделевым и А. Н. Дмитриевым [2] было предложено оценивать роли признаков в решении задачи распознавания как долю вхождения их в $\mathfrak{Z}(T)$. Эта величина p_j для признака x_j , называемая информационным весом, позволила им рассмотреть при специальной кодировке значений признаков линейный функционал $\sum_{j=1}^n p_j x_j$, ввести некоторые пороги d_i для T_i и по значению $\sum_{j=1}^n p_j x_j$ и d_i определять близость набора значений признаков $\alpha = (a_1, a_2, \dots, a_n)$ к некоторой из матриц T_i . На этом пути ими было предложено решение ряда задач по оценке месторождений полезных ископаемых.

Позже В. И. Переяславским [3] был исследован вопрос, когда линейный функционал $\sum_{j=1}^n p_j x_j$ доставляет верное решение задачи распознавания.

Сюда примыкают рассуждения А. Шайбеа [4], посвященные выяснению возможностей линейных функционалов в решении задачи распознавания в общем случае.

Другое развитие идеи использования тестов было осуществлено автором. Как отмечалось, тест σ может выступать в качестве «эксперта» для определения по набору α состояния q объекта A . В случае, когда $\sigma \in \mathfrak{Z}(T')$, он по α из T уже, вообще говоря, не решает задачи принадлежности α некоторому T_i , поскольку $\mathfrak{Z}(T')$ может не совпадать с $\mathfrak{Z}(T)$. Более того, он может отнести α к другой матрице T_j , или вообще отказаться от принятия решения, когда α_σ не входит ни в одну из матриц $T'_{i,\sigma}$. Таким образом, возникает необходимость подвергнуть анализу набор α с помощью всего доступного нам множества $\mathfrak{Z}(T')$. Используя каждый тест σ из $\mathfrak{Z}(T')$ для α , получаем вектор «голосов» $\chi(\alpha) = (k_1, \dots, k_m, k_{m+1})$, где k_i — число тестов, высказавшихся за T_i при $i = 1, 2, \dots, m$, а k_{m+1} — число «воздержавшихся» голосов.

Можно считать, что в векторе χ координаты пронормированы, то есть поделены на число тестов в $\mathfrak{Z}(T')$. Тогда k_l интерпретируем как меру выраженности свойства α принадлежать T_l при $l \leq m$, а k_{m+1} — не принадлежать T' ; здесь следует полагать, что $\mathfrak{Z}(T')$ и $\mathfrak{Z}(T)$ отличаются достаточно «мало».

Функционал χ был успешно опробован в решении ряда прикладных задач, но вместе с тем обнаружилось и определенные трудности в его использовании. Главными из них являлись сложностные факторы и слабое согласование численных результатов с реальностью для отдельных задач распознавания. Это привело к необходимости отхода от голосования по всему множеству тестов и к замене его специальными семействами. К их числу относятся тупиковые, короткие, минимальные тесты и их ограничения.

Тупиковый тест представляет собой тест, у которого любое собственное подмножество признаков не образует тест.

Короткий тест по числу признаков близок к логарифму числа признаков в T .

Минимальный тест имеет наименьшее число признаков из всех возможных.

Возникающие функционалы распознавания, соответствующие голосованию по этим семействам, обозначаем через χ_{TT} , χ_{KT} и χ_{MT} , а сами семейства называем основными.

Выяснились огромные преимущества функционала χ_{KT} в решении задач распознавания.

С его помощью было решено большое число задач геологии, экономики, военного дела и других областей.

Были, например, установлены новые месторождения нефти, газа и олова в Сибири [5], оценены перспективы развития конкретных экономических районов, проводился анализ текущей военно-политической ситуации, устанавливался диагноз заболевания и, соответственно, оптимальный режим лечения для него [6].

Приблизительно в то же время идея голосования была использована Ю. И. Журавлевым и его учениками при голосовании по подмножествам признаков, состоящих из заданного числа элементов. В частности, предложена оптимизация выбора конкретного значения r , при котором распознавание оказывается наилучшим [7].

Особая значимость функционалов χ , χ_{TT} , χ_{KT} и χ_{MT} в решении задач распознавания, подтвержденная практикой, создала предпосылки для разработки тестовой теории распознавания. Ее создание предполагало решение следующих задач.

- 1) Получение оценок для числа основных видов тестов матрицы T , имеющей заданные параметры ее подматриц T_i .
- 2) Нахождение точных и приближенных алгоритмов построения основных семейств тестов.
- 3) Выяснение того, когда для T и T' заданные виды их основных семейств отличаются на заданную долю.
- 4) Решение обратной задачи для 3.
- 5) Выделение того семейства из основных, которое для T' лучше остальных решает задачу распознавания.
- 6) Нахождение быстрых алгоритмов вычисления функционалов распознавания для T , соответствующих основным семействам.
- 7) Выяснение роли отдельных признаков в решении задачи распознавания и определение корреляции между ними для T' .
- 8) Решение задач 1–7 при заданном графе сравнения для T'_i в T' с соответствующим уточнением понятий видов тестов.
- 9) Решение задач 1–8 для почти всех матриц T' .

В решении задач 1 и 2 в случае, когда в T' каждая подматрица T'_i состояла из одной строки, первые результаты были получены Слепян [8] и Носковым [9], оценивших сверху и снизу число тестов и тупиковых тестов в таких матрицах.

Затем в предположении, что в T признаков «много больше», чем строк, Е. В. Дюковой [10] была разработана специальная комбинаторная техника, с помощью которой ею были найдены асимптотики тестов и тупиковых тестов для почти всех таблиц T' с заданными соотношениями параметров ее размерностей; эта техника ей позволила синтезировать оптимальные детерминированные и стохастические алгоритмы построения тестов и тупиковых тестов, а также решить задачу поведения весов признаков и значений их корреляций. Таким образом, Е. В. Дюковой удалось частично решить задачи 1, 2, 6, 7, 9.

Работа Е. В. Дюковой была прорывной по своему значению в создании теории тестового распознавания.

Достижения Е. В. Дюковой затем были развиты и существенно усилены А. Е. Андреевым [11]. Им была разработана новая специальная техника, позволившая:

- найти асимптотические поведения для числа тестов для произвольных размерностных характеристик матриц;
- построить соответствующие оптимальные алгоритмы детерминированного и стохастического типа для нахождения указанных тестов;
- частично решить задачу 4, указав возможные доопределения матриц T'_i таким образом, чтобы множества $\mathfrak{Z}(T')$ и $\mathfrak{Z}(T)$ совпадали;
- найти длины минимальных тестов и их число;
- выявить веса признаков и корреляции между ними;
- распространить результаты в решении задач 1, 2, 6, 7 на случай произвольного графа сравнимости в таблице T для почти всех таблиц.

Таким образом, А. Е. Андрееву удалось, решив в определенной мере задачи 1, 2, 3, 6, 7, 8 и 9, кардинально расширить базу знаний в теории тестов. В то же время из его результатов и частично из результатов Е. В. Дюковой вытекало, что эффективность распознавания процедур, основанных на тестах, существенно зависит от вида семейств тестов из основных, по которым идет распознавание.

Этому вопросу были посвящены исследования А. А. Кибкало [12]. Оказалось, что в качестве оптимального множества «голосующих» тестов следует выбирать так называемые «короткие» тесты. Если таблица T имеет m строк, то под коротким тестом для T' понимается величина «близкая» к $\log m - \log \log m$.

А. А. Кибкало для случая, когда T' состоит из двух подматриц, установлено, что:

- вес признака по тупиковым тестам уменьшается с увеличением доли различаемых строк из T'_1 и T'_2 ;

- вес признака по тестам не зависит от доли различаемых строк из T'_1 и T'_2 и почти всегда равен $\frac{1}{2}$;
- вес признака по коротким тестам растет с увеличением доли различаемых строк из T'_1 и T'_2 ;
- вес признака, различающего больше половины пар строк по тестам, длина которых близка к минимальной, почти всегда равен единице.

Таким образом, содержательно становится ясным преимущество коротких тестов и «очень» коротких тестов, то есть близких к минимальным тестам перед остальными в выявлении признаков, лучше других отличающих подматрицы T'_1 и T'_2 . Он существенно продвинулся в решении задачи 7.

Далее была изучена ситуация, близкая к задачам 3 и 5. А. А. Кибкало установил, что в случае, когда в матрице T' возникают ситуации:

- множество почти всех тупиковых тестов перестает быть таковым после искажения;
- множество очень коротких тестов почти полностью меняется;
- множество коротких тестов практически не меняется.

Им же построен асимптотически оптимальный алгоритм перечисления коротких тестов, то есть решены для рассматриваемого случая задачи 2 и 9.

Все задачи 1–9 в общем случае относятся к числу переборных и потому возникающие для них алгоритмы при больших размерах матриц T' мало применимы. Возникает необходимость перехода к более простым, но менее точным алгоритмам их решения.

Исторически первый приближенный алгоритм решения задач 6 и 7 был предложен В. Е. Кузнецовым [13]. Этот алгоритм типа Монте-Карло использовался в процедурах решения практических задач распознавания. Затем возникли приближенные алгоритмы для этих задач у Е. В. Дюковой, А. Е. Андреева и А. А. Кибкало.

Позже М. В. Носовым [14] был предложен алгоритм вычисления χ , основанный на экспоненциально сложном анализе множества T' с последующим полиномиальным по сложности вычислении χ по набору α .

Восстанавливая хронологию, необходимо подчеркнуть, что первые задачи, которые удалось решить Ю. И. Журавлеву и его коллегам, имели матрицу T' с числом строк и столбцов, соответственно, до 20.

Работа В. Е. Кузнецова позволила решать подобные задачи с уже до ста строк и столбцов в T' .

Е. В. Дюкова расширила класс таких матриц до имеющих 200 столбцов и 50 строк.

А. Е. Андреев распространил эти возможности до матриц размерами в 200 строк и 200 столбцов.

А. А. Кибкало смог использовать уже матрицы размерами 500 строк и 500 столбцов.

Современные компьютеры позволяют еще на порядок и более увеличить размеры реально обрабатываемых матриц.

В рассмотренной ситуации в общем случае не предполагалось, что имеется некоторая связь между подматрицами T'_i матрицы T' . Специальный случай возникает при допущении такой связи. Здесь следует выделить рассмотрение такой модели.

Пусть A является техническим устройством с r входами и одним выходом. На входы Q поступают значения a из $E_k = \{0, 1, \dots, k-1\}$, а на выходе снимается значение $f(a_1, \dots, a_r)$ — некоторой функции k -значной логики. Считается, что на входах могут возникать ошибки определенного типа γ , меняющие набор (a_1, \dots, a_r) на (a'_1, \dots, a'_r) . Тогда в общем случае на подаваемом наборе (a_1, \dots, a_r) значение f меняется. Требуется указать некоторое число наборов, вычисляя f на которых, можно установить, была ли ошибка на входах A .

Эти наборы образуют тест и обычно требуется, чтобы он имел наименьшее число элементов. Далее допускается, что функция f может быть любой от фиксированного числа r переменных и ставится вопрос о том, каково минимально достаточное число $L_K(r, \gamma)$, такое что для любых из указанных функций f и ошибки типа γ найдется тест из $L_K(r, \gamma)$ элементов, который по f определяет, была ли на входах A , реализующего f , допущена ошибка γ .

Отметим здесь две группы основных результатов.

Первую составляют результаты Г. Р. Погосяна [15], который установил поведение $L_K(r, \gamma)$ для γ , соответствующих коротким замыка-

ниям, слияниям, инверсиям и др. Ранее в этом направлении интересные продвижения были получены В. Н. Носковым [9].

Вторую составляют результаты О. А. Долотовой [16], которая для тех же типов ошибок нашла поведение функции для $L_2^C(r, \gamma)$, аналогичной функции $L_C(r, \gamma)$, но соотношенной с заданным классом Поста C функций алгебры логики. К этим группам результатов примыкают подобные рассуждения для почти всех функций.

В целом же тестирование технических устройств привлекает большое число исследователей, которыми накоплен большой опыт, заслуживающий отдельного обсуждения. Упомянем здесь лишь итоговые изложения в [17].

Статья, помимо введения, содержит пять параграфов.

В первом параграфе приводятся тестовые результаты для матриц T' с малым числом строк.

Во втором параграфе излагаются результаты по теста в общем случае при произвольном графе сравнения подматрицы T' .

В третьем параграфе содержатся основные факты по тестированию схем, реализующих функции k -значной логики.

В пятом параграфе приводятся результаты по тестированию схем, реализующих функции из классов Поста.

1. Тесты для матриц с малым числом строк

Здесь мы изложим основные результаты Е. В. Дюковой по задачам 6 и 7.

Пусть матрица T имеет m строк и n столбцов, а ее элементы a_{ij} входят в $E_k = \{0, 1, \dots, k-1\}$, $k \geq 2$. Пусть $m = \sum_{i=1}^s m_i$, где m_i — натуральные и $m_i \leq 1$; \tilde{m} есть последовательность m_1, m_2, \dots, m_s . Пусть $M_{\tilde{m}, n, s}^{(k)}$ — класс всех таких матриц T' , состоящих из s подматриц T'_i с m_i строками, следующими в T друг за другом с ростом номеров этих подматриц.

При $l \in \{1, 2, \dots, s\}$ полагаем $m''(l) = \sum_{i=1}^l m_i$, $m'(l) = m''(l) - m_l + 1$.

Считаем, что в матрице из $M_{\tilde{m}, n, s}^{(k)}$ подматрица T'_l содержит строки с

номера $m'(l), m'(l+1), \dots, m''(l)$. Пусть $\mathfrak{S}_t(T')$ — множество всех тупиковых тестов для T' . Для строк $\alpha = (a_{j_1}, a_{j_2}, \dots, a_{j_n})$ и $\alpha' = (a_{j'_1}, a_{j'_2}, \dots, a_{j'_n})$ из T' вводим сложение по модулю два \oplus , полагая

$$\alpha \oplus \alpha' = (a_{j_1} \oplus a_{j'_1}, a_{j_2} \oplus a_{j'_2}, \dots, a_{j_n} \oplus a_{j'_n}). \quad (1)$$

Пусть $h = \sum_{l=1}^{s-1} \sum_{t=1}^{s-l} m_l m_{l+t}$, составим матрицу $L_{T'}$ с элементами b_{uv} , $u \in \{1, 2, \dots, h\}$, $v \in \{1, 2, \dots, n\}$ из всех строк $\alpha \oplus \alpha'$ таких, что строки α и α' входят в разные подматрицы T'_i и $T'_{i'}$, соответственно, причем $i < i'$. Полагаем строки в $L_{T'}$ упорядоченными так, что если $\alpha, \alpha', \alpha'', \alpha'''$ имеют номера в T суть j_1, j_2, l_1, l_2 , соответственно, то если номер строки $\alpha \oplus \alpha'$ меньше номера строки $\alpha'' \oplus \alpha'''$, то либо $j_1 < l_1$, либо $j_2 < l_2$.

Матрица $L_{T'}$ называется матрицей сравнения для T' . Каждому элементу b_{uv} из $L_{T'}$ присваивается номер $N(u, v) = (v-1) \cdot h + u$.

Рассмотрим класс \mathcal{A} алгоритмов, строящих все тупиковые тесты для T' . Алгоритм a из \mathcal{A} сводит эту задачу к некоторой задаче следующего вида.

Имеется множество $\mathfrak{R} = \{R_1, R_2, \dots, R_g\}$, где R_W — некоторый набор элементов из $L_{T'}$, $W = \{1, 2, \dots, g\}$. Требуется построить все такие подмножества R , для каждого из которых выполнен ряд свойств B_1, B_2, \dots, B_p и $p \geq 2$.

Искомую совокупность обозначим через $Q_{\mathfrak{R}}$. Предполагаем, что между $Q_{\mathfrak{R}}$ и $\mathfrak{S}_t(T')$ существует взаимно-однозначное соответствие, так что построение $Q_{\mathfrak{R}}$ приводит к построению $\mathfrak{S}_t(T')$.

Пусть $Q^{(t)}$, $t \in \{1, 2, \dots, p\}$ — совокупность всех таких подмножеств множества \mathfrak{R} , для каждого из которых выполнены свойства $B_1, B_2, \dots, B_{t-1}, B_{t+1}, \dots, B_p$. На систему свойств $\{B_1, B_2, \dots, B_p\}$ наложим условие: всегда $Q^{(t)} \supset Q_{\mathfrak{R}}$.

Пример 1. Пусть $g = u$ и элемент $R_W \in \mathfrak{R}$ является столбцом матрицы $L_{T'}$ под номером W . Для наборов столбцов матрицы L_T рассмотрим такие свойства B_1 и B_2 :

- набор столбцов H матрицы $L_{T'}$ обладает свойством B_1 , если для любой строки из $L_{T'}$ в H найдется столбец, пересечение которого с этой строкой дает 1;

- набор столбцов H матрицы $L_{T'}$ обладает свойством B_2 , если из условия $H' \subset H$ следует, что H' не обладает свойством B_1 .

Требуется построить все такие наборы столбцов из $L_{T'}$, каждый из которых обладает свойствами B_1 и B_2 .

Решение этой задачи приводит к построению $\mathfrak{S}_t(T')$. В самом деле набор σ номеров столбцов j_1, j_2, \dots, j_r матрицы T' является тупиковым тестом точно тогда, когда набор столбцов матрицы $L_{T'}$, с указанными номерами обладает свойствами B_1 и B_2 .

Пример 2. Два различных единичных элемента b_{uv} и $b_{u'v'}$ матрицы $L_{T'}$ считаются совместимыми, если $b_{uv} = b_{u'v'} = 0$. Множество D из r единичных элементов матрицы $L_{T'}$ называем совместимыми, если выполнено хотя бы одно из условий:

- 1) $r = 1$;
- 2) $r > 1$ и любые два элемента в D совместимы.

Множество всех совместимых множеств в $L_{T'}$ обозначим через $S(L_{T'})$.

Пусть $g = h \cdot n$, элемент R_W множества \mathfrak{R} является элементом матрицы $L_{T'}$ с номером $W \in \{1, 2, \dots, g\}$ и $\Omega(D)$ — набор тех столбцов из $L_{T'}$, в которых расположены элементы множества $D \subseteq R$. Говорят, что D обладает свойством B'_1 , если $D \in S(L_{T'})$. Если $D = \{b_{u_1v_1}, b_{u_2v_2}, \dots, b_{u_rv_r}\}$, D обладает свойством B'_2 , если выполнены условия:

- 1) $\Omega(D)$ обладает свойством B_1 ;
- 2) $\{b_{u_1v_1}, b_{u_2v_2}, \dots, b_{u_rv_r}\} \in S(L_{T'})$, то $u_t \leq p_t$ при $t \in \{1, 2, \dots, r\}$.

Множество $D \subseteq \mathfrak{R}$ называется правильным в $L_{T'}$, если D обладает свойствами B'_1 и B'_2 . Пусть $\mathcal{B}(L_{T'})$ — множество всех правильных множеств в $L_{T'}$. Построим $\mathcal{B}(L_{T'})$.

Тупиковый тест таблицы T' $\sigma = (j_1, j_2, \dots, j_r)$ и множество $D \in \mathcal{B}(L_{T'})$ считаем эквивалентными, если $\Omega(D)$ состоит из столбцов матрицы $L_{T'}$ с номерами j_1, j_2, \dots, j_r .

Можно показать, что между $\mathcal{B}(L_{T'})$ и $\mathfrak{S}_t(T')$ имеется взаимно-однозначное соответствие, при котором соответствующие элементы

эквиваленты. Следовательно, построение $\mathcal{B}(L_{T'})$ приводит к построению $\mathfrak{S}_t(T')$.

При построении множества $Q_{\mathfrak{R}}$ алгоритм \mathcal{A} поступает так. Из системы свойств $\mathcal{B} = \{B_1, B_2, \dots, B_p\}$ выделяется некоторая подсистема $\mathcal{B}' = \{B_{i_1}, B_{i_2}, \dots, B_{i_j}\}$ такая, что $\mathcal{B}' \subset \mathcal{B}$. Пусть Q_{i_1, i_2, \dots, i_l} — множество всех таких подмножеств из \mathfrak{R} , для каждого из которых выполнены свойства $B_{i_1}, B_{i_2}, \dots, B_{i_l}$. Работа алгоритма \mathcal{A} состоит в построении множества Q_{i_1, i_2, \dots, i_l} и выделению из него подмножества $Q_{\mathfrak{R}}$, при этом на каждом его шаге строится некоторый элемент из Q_{i_1, i_2, \dots, i_l} и для него проверяется выполнение свойств B_t , где $t \in \{1, 2, \dots, p\} \setminus \{i_1, i_2, \dots, i_l\}$.

Множество Q_{i_1, i_2, \dots, i_l} назовем погружением для алгоритма \mathcal{A} .

Пусть $|M|$ — мощность множества M , $M(a, T') = |Q_{i_1, i_2, \dots, i_l}|$, $N_{m, n} \subseteq M_{m, n, s}^{(k)}$.

Алгоритм \mathcal{A} асимптотически оптимальный в $N_{m, n}$, если для почти всех подматриц T' из $N_{m, n}$ выполнено

$$\mu(\mathcal{A}, T') \sim |\mathfrak{S}_t(T')| \quad (2)$$

при $m, n \rightarrow \infty$.

Здесь будет предполагаться, что число столбцов в матрице T' «много» больше числа строк.

Нетрудно убедиться, что примером алгоритма для построения $\mathfrak{S}_t(T')$, не являющимся асимптотически оптимальным, является процедура из [10], которую обозначаем через \mathcal{A}_1 .

Алгоритм \mathcal{A}_1 сводит задачу построения $\mathfrak{S}_t(T')$ к задаче, описанной в примере 1 и для построения соответствующего ему погружения использует свойство B_1 . Величина $\mu(\mathcal{A}, T')$, равная числу тестов в T' , почти всегда по порядку больше величины $|\mathfrak{S}_t(T')|$.

Сведение задачи построения $\mathfrak{S}_t(T')$ к задаче из примера 1 позволяет построить асимптотически оптимальный по сложности алгоритм. Этот алгоритм \mathcal{A}_1 для построения погружения использует свойство B'_1 и для него $\mu(\mathcal{A}, T') = |S(T')|$. Следовательно, вопрос об асимптотической оптимальности \mathcal{A}_1 сводится к оценке величины $|S(T')|$ и $|\mathfrak{S}_t(T')|$.

Справедливо утверждение.

Теорема 1. Если $\alpha > 1$, $\beta < \frac{1}{2}$, $(m_1, m_2)^\alpha \leq n \leq k^{(m_1, m_2)^\beta}$, $\varphi_k = (a, b)$, где $a = \frac{1}{2} \log_k(m_1 m_2 n) - \frac{1}{2} \log_k \log_k(m_1 m_2 n) - \log_k \log_k \log_k n$, $b = \frac{1}{2} \log_k(m_1 m_2 n) - \frac{1}{2} \log_k \log_k(m_1 m_2 n) + \log_k \log_k \log_k n$, то при $m, n \rightarrow \infty$ выполнено для почти всех T из M_m

$$|S(L_{T'})| \sim |\mathfrak{S}_t(T')| \sim \sum_{r \in \varphi_k} \frac{[m_1 m_2 n (k-1)]^r}{r! k^{k^2}}. \quad (3)$$

Отсюда следует, что при $(m_1 m_2)^\alpha \leq n \leq k^{(m_1 m_2)^\beta}$, $\alpha > 1$, $\beta < \frac{1}{2}$ алгоритм \mathcal{A}_1 является асимптотически оптимальным в $F_{m_1, m_2, n}^{(k)}$.

При доказательстве теоремы 1 множество матриц $F_{m_1, m_2, n}^{(k)}$ рассматривается как пространство элементарных событий, в котором каждое событие $T' \in F_{m_1, m_2, n}^{(k)}$ происходит с вероятностью $\frac{1}{|F_{m_1, m_2, n}^{(k)}|}$. Оцениваются вероятностные характеристики распределения случайных величин $|S(L_{T'})|$ и $|\mathfrak{S}_t(T')|$. Показывается, что в условиях теоремы 1 математические ожидания величин $|S(L_{T'})|$ и $|\mathfrak{S}_t(T')|$ асимптотически совпадают, а, с другой стороны, каждая из величин $|S(L_{T'})|$ и $|\mathfrak{S}_t(T')|$ почти всегда асимптотически совпадает со своим математическим ожиданием.

Следует отметить, что асимптотика величины $|\mathfrak{S}_t(T')|$ для почти всех матриц T' из $M_{m, n, s}^{(2)}$ в случае, когда $s = m$, найдена в работе В. А. Слепян и В. Н. Носкова [8], [9], где при получении оценки для $|\mathfrak{S}_t(T')|$ так же, как и в доказательстве теоремы 1, используется то, что почти всегда величина $|\mathfrak{S}_t(T')|$ асимптотически совпадает со своим средним значением.

Наряду с детерминированной процедурой построения множества $\mathfrak{S}_t(T')$ Е. В. Дюковой построены стохастические алгоритмы, основанные на алгоритме \mathcal{A}_1 .

При стохастическом подходе используется не все множество типовых тестов матрицы T' , а лишь случайная выборка из него, анализ которой дает возможность приближенно решать задачи быстрого распознавания, выявления роли отдельных признаков в задаче распознавания и корреляции между ними, то есть, соответственно, задачи 6 и 7, а также оценивать при этом возможную ошибку.

Так как объем выборки существенно меньше множества всех типовых тестов, то при этом удается значительно увеличить размеры обрабатываемых таблиц.

В упомянутом ранее алгоритме В. Е. Кузнецова [13] построение тупиковых тестов для T' из $M_{m,n,s}^{(k)}$ осуществляется по следующей схеме.

Задаются целые числа q , $1 \leq q < n$ (число q выбирается особым образом и в сравнении с n достаточно мало). Пусть W_q — множество всех наборов вида $\{j_1, j_2, \dots, j_q\}$, где $j_t \in \{1, 2, \dots, n\}$ при $t \in \{1, 2, \dots, q\}$ и $j_1 < j_2 < \dots < j_q$. Случайным образом выбираются наборы из W_q . Каждый такой набор $W = \{j_1, j_2, \dots, j_q\}$ определяет подматрицу T'_W в T' , образованную столбцами с номерами j_1, j_2, \dots, j_q .

К матрице T'_W применяется алгоритм \mathcal{A}_2 , который строит все тупиковые тесты для T_W . Выбор некоторой совокупности случайных наборов из W_q и приводит к построению случайной выборки из $\mathfrak{S}_t(T')$.

Пусть $i \in \{1, 2, \dots, h\}$, где h — число строк в матрице сравнения $L_{T'}$ для T'_u , U_i — совокупность всех наборов вида $\{l_1, l_2, \dots, l_i\}$, $l_t \in \{1, 2, \dots, h\}$ при $t \in \{1, 2, \dots, i\}$ и $l_1 < l_2 < \dots < l_i$.

Пусть $u \in U_i$. Тупиковый тест для T' такой, что $\sigma = \{j_1, j_2, \dots, j_r\}$, называется u -тестом, если в u можно указать числа l_1, l_2, \dots, l_r такие, что элементы с номерами $N[l_1, j_1], N[l_2, j_2], \dots, N[l_r, j_r]$ образуют совместимое множество в $L_{T'}$. Совокупность всех u -тестов матрицы T' обозначаем через $\mathfrak{S}(T', u)$.

Алгоритм \mathcal{A}_1 быстро работает на таких таблицах, у которых число строк много меньше числа столбцов. Поэтому предлагаемый стохастический алгоритм построения тупиковых тестов для T' использует не случайные наборы из W_q , а случайные наборы из U_i . Каждый такой набор $u = \{l_1, l_2, \dots, l_i\}$ определяет подматрицу $L_{T'}(u)$ матрицы $L_{T'}$, образованную строками с номерами l_1, l_2, \dots, l_i . Матрица $L_{T'}(u)$ обрабатывается при помощи алгоритма \mathcal{A}_1 , который строит $\mathfrak{S}(T', u)$. Обработка некоторой совокупности случайных наборов из U_i приводит к построению случайной выборки из $\mathfrak{S}_t(T')$.

Эта схема в деталях отличается от схемы В. Е. Кузнецова, в то время как при нахождении меры важности признака оказывается, что при определенных условиях оба подхода почти всегда эквивалентны детерминированному подходу.

Точнее. Пусть $\mathfrak{S}_{t,j}(T')$ — множество всех тупиковых тестов, содержащих признак x_j , а $\mathfrak{S}_j(T', u)$ — множество тупиковых тестов в $\mathfrak{S}(T', u)$, содержащих признак x_j , тогда наряду с информационным весом введем величины:

$$p(x_j) = \frac{|\mathfrak{S}_{t,j}(T')|}{|\mathfrak{S}(T')|}, \quad (4)$$

$$p^i(x_j) = \frac{1}{C_n^i} \sum_{u \in U_i} \frac{|\mathfrak{S}_i(T', u)|}{|\mathfrak{S}(T', u)|}, \quad (5)$$

$$\bar{p}^i(x_j) = \frac{\sum_{u \in U_i} |\mathfrak{S}_i(T', u)|}{\sum_{u \in U_i} |\mathfrak{S}(T', u)|}, \quad (6)$$

$$\bar{p}^i(x_j, u) = \frac{|\mathfrak{S}_i(T', u)|}{|\mathfrak{S}(T', u)|}, \quad (7)$$

где $u \in U_i$.

Теорема 2. Если $u \in U_i$, $n \leq k^{(m_1 m_2)^\beta}$, $\beta < \frac{1}{3}$, и при $m, n \rightarrow \infty$ имеет место:

$$\frac{\log_k(m_1 m_2)}{\log_k n} \rightarrow 0, \quad \frac{\log^3 n}{i} \rightarrow 0, \quad (8)$$

то почти для всех матриц T' из $M_{\bar{m}, n, s}^{(k)}$ при $n \rightarrow \infty$ справедливо

$$p(x_j) \sim p^i(x_j) \sim \bar{p}^i(x_j) \sim \frac{\log_k n}{2n}. \quad (9)$$

Эта теорема устанавливает асимптотическую эквивалентность стохастического и детерминированного подходов в тестовых алгоритмах распознавания, использующих в функционалах принятия решений веса признаков.

2. Тесты для матриц с заданным графом сравнения

Здесь мы изложим основные результаты А. Е. Андреева по задачам 1, 2, 6, 8, 9, которые существенно усиливают результаты Е. В. Дюковой.

Введенные ранее понятия теста, тупикового теста и другие в более общей ситуации, которая рассматривается здесь, нуждаются в уточнениях. С этой целью введем необходимый формализм.

Для множеств A и B полагаем $A^B = \{f | f : B \rightarrow A\}$. Если $f \in A^B$ и $C \subseteq B$, то $f(C) = \{b | b = f(a), a \in C\}$; а если $a \in A$ и $C \subseteq A$, то $f^{-1}(a) = \{b | f(b) = a\}$ и $f^{-1}(C) = \bigcup_{a \in C} f^{-1}(a)$.

Пусть $\langle A^B \rangle = \{f | f \in A^B, f(b) = a\}$. Если $f \in A^B, g \in B^C$, то $f \circ g \in A^C$ и $(f \circ g)(x) = f(g(x))$. Пусть $A \otimes B$ — множество упорядоченных пар (a, b) , где $a \in A, b \in B$; $A \times B$ — множество неупорядоченных пар $\{(a, b)\}$, причем $a \neq b$. Обозначим через $A^{(2)}$ множество $A \times A$. Если A — некоторое конечное множество, то через 1^A и 0^A обозначим, соответственно, константы 1 и 0 из E_2^A ; если $x, y \in A$, то полагаем $1_x^A(y) = 1$, если $x \neq y$, и $1_x^A(y) = 0$, если $x = y$; считаем, что $0_x^A(y) = 1_x^A \oplus 1^A, x \in A$. Положим $\tilde{\varepsilon}^A = \{1_x^A | x \in A\}$, $\varepsilon^A = \tilde{\varepsilon}^A \cup \{1^h\}$. Знаки \leq и $<$ для функций из E^A используем в обычном смысле для указания поэлементной упорядоченности на этом множестве. Если $a, b \in E^A$, то полагаем $a * b = a \oplus b \oplus 1^A$; $|a| = |a^{-1}(1)|$; $\rho(a, b) = |a \oplus b| = |A| - |a * b|$.

При $m \geq 1$ полагаем $L_m = \{1, 2, \dots, m\}$. отождествим наборы из E_2^n , n -мерного единичного куба, с отображениями из F^{L_m} . Другими словами, i -я координата набора a является значением отображения $a(i)$, и можем упорядоченные наборы длины n элементов из A рассматривать как отображения из A^{L_m} . Пусть $E^* = \bigcup_{n=1}^{\infty} E^n$, $E_r^n = \{x | x \in E^n, |x| = r\}$. Положим $1^n = 1^{L_n}, 0^n = 0^{L_n}; 1_i^n = 1_i^{L_n}, 0_i^n = 0_i^{L_n}, i \in L_n; \tilde{\varepsilon}^n = \tilde{\varepsilon}^{L_n}, \varepsilon^n = \varepsilon^{L_n}$. Пусть $a \in E^n$, определим отображение η_a из $L_n^{L_{|a|}}$, положив $\eta_a(i) = k$, если $\sum_{j=1}^{k-1} a(j) < i$, где $i = \sum_{j=1}^k a(j)$.

Эта функция интерпретируется так. Пусть $a, b \in E^n$, тогда если выписать набор b и вычеркнуть координаты, соответствующие нулям набора a , то получим запись набора $b \circ \eta_a$. Для a из E_2^n определяем отображение P_r^a из $(E^{|a|})^{E^n}$, полагая, что $P_r^a(b) = b \circ \eta_a$.

Введем ряд обозначений, связанных с графами. Пусть $V(G)$ и $X(G)$ — множества вершин и, соответственно, ребер графа G . Предполагаем, что $X(G) \subseteq V(G)^{(2)}$. Пусть $\gamma_G(x)$ — степень вершины x , $\Delta(G)$ — максимальная степень вершины, $\chi(G)$ — хроматическое чис-

ло, $d(r, G)$ — число остовных r -реберных лесов. Через $k(G)$ обозначим число компонент связности, а через $k^*(G)$ — число нетривиальных компонент, то есть содержащих не менее двух вершин.

Пусть $p(G) = |V(G)|$, $q(G) = |X(G)|$, $p^*(G) = p(G) - k(G) + 1$. Если U — конечное множество и $M = \{V_1, V_2, \dots, V_k\}$ — система его подмножеств, то M называется покрытием для U , если $\bigcup_{i=1}^k V_i = U$.

Через $G(M)$ обозначим граф с множеством вершин U и множеством ребер $U^{(2)} \setminus \left(\bigcup_{i=1}^k V_i^{(2)} \right)$.

Назовем бинарной U -матрицей, имеющей n столбцов, отображение из $(E_2^n)^U$.

Если M — покрытие для U , то через $\Psi_{M,n}$ обозначим множество пар (T, M) , где $T \in (E_2^n)^U$; если $V(G) = U$, то через $\Psi_{G,r}$ обозначим множество пар (T, G) таких, что $T \in (E^n)^U$. Обобщим ранее введенное понятие теста до понятия G -теста, но для краткости сохраним за ним термин тест.

Пусть заданы конечное множество U , его покрытие $M = \{V_1, V_2, \dots, V_k\}$ и граф G с множеством вершин U . Тогда набор a из E_2^n назовем:

- 1) тестом пары (T, M) из $\Psi_{M,n}$, если для любых a_1, a_2 из U таких, что $T(a_1) * T(a_2) \geq a$, найдется i из L_k , что $a_1, a_2 \in V_i$;
- 2) тестом пары (T, G) из $\Psi_{G,n}$, если ни для каких смежных вершин a_1, a_2 графа G не выполнено $T(a_1) * T(a_2) \geq a$;
- 3) тестом матрицы T из $(E^n)^U$, если ни для какого b из U не выполнено $T(b) \geq a$.

Тест a пары (T, M) (пары (T, G) , матрицы T) называется тупиковым, если никакой выбор b из E^n такой, что $b \leq a$, не является тестом пары (T, M) (пары (T, G) , матрицы T).

Для T из $(E^n)^U$ определяем $U^{(2)}$ — матрицу $T^{(2)}$, положив $T^{(2)}(\{a_1, a_2\}) = T(a_1) * T(a_2)$. Для (T, G) из $\Psi_{G,n}$ полагаем $T_G = T^{(2)}|_{X(G)}$ и называем ее матрицей сравнения этой пары; для пары (T, M) из $\Psi_{M,n}$, где T есть U -матрица, а M — покрытие множества U , положим $T_M = T_{G(M)}$ и назовем ее матрицей сравнения пары (T, M) . Непосредственно из определения вытекает следующее утверждение.

Теорема 3. Следующие условия эквивалентны:

- 1) набор a является тестом (тупиковым тестом) пары (T, M) , где T является U -матрицей, а M — покрытием U ;
- 2) набор a является тестом (тупиковым тестом) пары $(T, G(M))$;
- 3) набор a является тестом (тупиковым тестом) таблицы T_M .

Пусть N , R и R_+ , соответственно, множества натуральных, действительных и положительных действительных чисел, а \mathcal{M} — множество таких непрерывных отображений α из $R_+^{R_+}$, что $\lim_{x \rightarrow \infty} \alpha(x) = 0$, а $\mathcal{B} = \{D \mid D = \frac{1}{\alpha}, \alpha \in \mathcal{M}\}$.

Знаки \leq_n° , \geq_n° , $=_n^\circ$ используются для обозначения неравенств и равенств, которые выполняются, начиная с некоторого n_0 . Точнее, если существует α из \mathcal{M} такое, что $B \leq_n^\circ (1 + \alpha(n))A$, то пишем $B \lesssim_n A$ (при $n \rightarrow \infty$ асимптотически не превосходит); если $B \lesssim_n A$ и $A \lesssim_n B$, то пишем $A \sim_n B$ (при $n \rightarrow \infty$ асимптотически равны). Если выполнено $c_1 A \leq B \leq c_2 A$ (либо $c_1 A \leq_n^\circ B \leq_n^\circ c_2 A$), где $c_1, c_2 \in R_+$, то пишем $B \asymp A$ (либо $B \asymp_n A$). Если для $\alpha \in \mathcal{M}$ верно, что $A \leq_n^\circ \alpha(n)B$, то пишем $A \ll_n B$.

Говорим, что при некоторых ограничениях на пары (G, n) свойство выполнено для почти всех при $n \rightarrow \infty$ пар (T, G) из $\Psi_{(G, n)}$, если существует α из \mathcal{M} , что доля пар (T, G) из $\Psi_{(G, n)}$ (где (G, n) удовлетворяет упомянутым ограничениям), для которых это свойство не выполнено, не превосходит $\alpha(n)$. Если это свойство — асимптотическое равенство функций $A_{G, n}$ и $B_{G, n}$, определенных на $\Psi_{(G, n)}$, то пишем $A_{G, n} \sim_n^{\text{п.в.}} B_{G, n}$ (это означает, что существует α и β из \mathcal{M} , что доля пар (T, G) из $\Psi_{(G, n)}$, для которых не выполнено $A_{G, n}(T, G)(1 - \beta(n)) \leq B_{G, n}(T, G) \leq (1 + \beta(n))A_{G, n}(T, G)$ не превосходит $\alpha(n)$). Аналогично определяются знаки $\leq_n^{\text{п.в.}}$, $=_n^{\text{п.в.}}$, $\geq_n^{\text{п.в.}}$ и т. д.

Значения функций $\varphi_{G, n}^{T, T}$ и $\varphi_{G, n, r}^{T, T}$ на матрице T равны, соответственно, числу тупиковых тестов и числу тупиковых тестов длины r пары (T, G) . Далее через $\log x$ обозначаем $\log_2 x$. Пусть $\chi(x) = (\log x - \log \log x)/2$ и $\chi(x) = (\log x - \log \log x)/2$, $H(m, r) = e^{-m \cdot 2^{-r}}(1 - e^{-mr \cdot 2^{-r}})^r$. Положим для действительного x , натуральных m и n таких, что $0 \leq x \leq n$, $\hat{H}(n, m, r) = \binom{n}{x} H(m, r)$. Для $0 \leq x \leq n$ полагаем

$\binom{n}{x} = \frac{\Gamma(n+1)}{\Gamma(x+1)\Gamma(n-x+1)}$, где Γ — гамма-функция, в противном случае считаем, что $\binom{n}{x} = 0$. Если существует, и притом единственное, относительно x , решение уравнения $\hat{H}(n, m, x+1) = \hat{H}(n, m, x-1)$ из интервала $(1, n-1)$, то полагаем $r(n, m)$ равным этому решению, иначе полагаем $r(n, m) = 1$.

Пусть $\varepsilon \in (0, \frac{1}{16})$ и D из \mathcal{B} , $D(n) \leq_n^\circ \log \log \log n$. Определим функцию $\Phi_D(G, n)$ так:

- 1) $\binom{n}{p^*(G)-1} d(p^*(G)-1, G) \cdot (p^*(G)-1)! 2^{-(p^*(G)-1)^2}$,
если $q(G) < D(n) \log^2 n$, $P^*(G) - 1 \leq \chi(n) - D(n)$;
- 2) $]\min(P^*(G) - 1, \chi(n \cdot q(G)) + D(n))$],
 $\sum \binom{n}{r} d(r, D) r! 2^{-r^2} r = [\chi(n) - D(n)]$,
если $q(G) \geq D(n) \log^2 n$.

Положим:

- 1) $\hat{r}_D^1(G, n) = \hat{r}_D^2(G, n) = p^*(G) - 1$,
если $q(G) < D(n) \log^2 n$, $p^*(G) - 1 \leq \chi(n) - D(n)$;
- 2) $\hat{r}_D^1(G, n) = [\chi(n) - D(n)]$,
 $\hat{r}_D^2(G, n) =]\min(P^*(G) - 1, \chi(n \cdot q(G)) + D(n))$],
если $q(G) < D(n) \log^2 n$, $p^*(G) - 1 > \chi(n) - D(n)$;
- 3) $\hat{r}_D^1(G, n) = [\chi(n)(nq(G) - D(n))]$,
 $\hat{r}_D^2(G, n) = [\chi(n)(nq(G) + D(n))]$,
если $q(G) \geq D(n) \log^2 n$.

Положим:

- 1) $r_{D,\varepsilon}^1(n, m) = [r(n, m) - D(n)]$,
 $r_{D,\varepsilon}^2(n, m) = [r(n, m) + D(n)]$,
если $m < n^{1+\varepsilon}$;
- 2) $r_{D,\varepsilon}^1(n, m) = r_{D,\varepsilon}^2(n, m) - 1 = [r(n, m)]$,
если $n^{1+\varepsilon} m \leq m < 2^{4 \ln^2 W}$;
- 3) $r_{D,\varepsilon}^1(n, m) = r_{D,\varepsilon}^2(n, m) - 1 = [\Theta(m)]$,
если $m \geq 2^{4 \ln^2 n}$.

Пусть

$$\Phi_{D,\varepsilon}^2(n, m) = \sum_{r=r_{D,\varepsilon}^1(n, m)}^{r_{D,\varepsilon}^2(n, m)} \binom{n}{r} H(m, r). \quad (10)$$

Положим

$$\begin{aligned} \tilde{r}_{D,\varepsilon}^1(G, n) &= \hat{r}_D^1(G, n) \\ \tilde{r}_{D,\varepsilon}^2(G, n) &= \hat{r}_D^2(G, n) \end{aligned}, \quad (11)$$

если $q(G) \leq n^{1-\varepsilon}$;

$$\begin{aligned} \tilde{r}_{D,\varepsilon}^1(G, n) &= r_{D,\varepsilon}^1(n, q(G)) \\ \tilde{r}_{D,\varepsilon}^2(G, n) &= r_{D,\varepsilon}^2(n, q(G)) \end{aligned}, \quad (12)$$

если $q(G) > n^{1-\varepsilon}$.

Пусть

$$\Phi_{D,\varepsilon}(G, n) = \begin{cases} \Phi_D^1(G, n), & \text{если } q(n) \leq n^{1-\varepsilon} \\ \Phi_{D,\varepsilon}^1(n, q(G)), & \text{если } q(n) > n^{1-\varepsilon} \end{cases} \quad (13)$$

Следующее утверждение дает описание асимптотического поведения числа тупиковых тестов для почти всех таблиц.

Теорема 4. *Существует $\varepsilon \in (0, \frac{1}{16})$, константы c_1, c_2 , $1 < c_1 < c_2$, D из B , $D(n) \leq_n^\circ \log \log \log n$, что выполнено:*

1) *если $1 \leq q(G) \leq 2^{n(1-\varepsilon)}$, то*

$$\varphi_{G,n}^{T.T.} \sim_n^{n.в.} \sum_{r=\tilde{r}_{D,\varepsilon}^1(G, n)}^{\tilde{r}_{D,\varepsilon}^2(G, n)} \varphi_{G,n,r}^{T.T.} \sim_n^{n.в.} \Phi_{D,\varepsilon}(G, n);$$

2) *если $q(G) \geq 2^n D(n)$, то*

$$\varphi_{G,n}^{T.T.} =_n^{n.в.} 0;$$

3) *не существует такой единственной функции f , определенной на множестве пар (G, n) , где G — граф и $n \in N$, что выполнено: если $c_1 2^n \leq q(G) \leq c_2 2^n$, то $\varphi_{G,n}^{T.T.} \sim_n^{n.в.} f(G, n)$ (либо $\varphi_{G,n}^{T.T.} =_n^{n.в.} f(G, n)$).*

Определим на $(E^n)^{V(G)}$ функцию $L_{G,n}^{\min}$, положив:

- 1) $L_{G,n}^{\min}(T) = n + 1$, если (T, G) не имеет тестов;
- 2) $L_{G,n}^{\min}(T) = \min\{r | \varphi_{B,n,r}^{T,T}(T) \neq 0\}$ — в остальных случаях.

Пусть $S(b, a) = \log b - \log \ln \binom{a}{\log b}$ и $S_1(b, a) = \log b - \log \ln \binom{a}{S(b,a)}$.

Пусть $\varepsilon \in (0, \frac{1}{16})$ и λ_ε из $(\frac{1}{2}, 1)$ является решением уравнения

$$(1 - \varepsilon)\lambda + \lambda \log l + (1 - \lambda) \log(1 - \lambda) = 0, \quad (14)$$

$$D \in B, \quad D(n) \leq_n^\circ \log \log n. \quad (15)$$

Положим функцию $L_{D,\varepsilon}(G, n)^1$ равной:

- 1) $] \log \chi(G)[$,
если $p^*(G) \leq (\ln n) \log \ln n$ или,
если $(\ln n) \log \ln n < p^*(G) \leq (\log n)^{D(n)}$ и $\chi(G) < 2((\ln n) \log \ln n)^{\frac{3}{4}}$;
- 2) $[S_1 \left(\binom{\chi(G)}{2}, n \right) + \varepsilon]$,
если $(\ln n) \log \ln n < p^*(G) \leq (\log n)^{D(n)}$ и $\chi(G) \geq 2((\ln n) \log \ln n)^{\frac{3}{4}}$;
- 3) $[S(\Delta(G), n) + \varepsilon]$,
если $(\log n)^{D(n)} < p^*(G)$, $q(G) < 2^{n-\lambda_\varepsilon}$ и

$$\Delta(G) > \frac{(1 - \varepsilon)q(G)(\ln q(G))}{\ln \binom{n}{\log q(G)}}; \quad (16)$$

- 4) $[S(q(G), n) + \varepsilon]$,
если $(\log n)^{D(n)} < p^*(G)$, $q(G) < 2^{n-\lambda_\varepsilon}$ и

$$\Delta(G) \leq \frac{(1 - \varepsilon)q(G)(\ln q(G))}{\ln \binom{n}{\log q(G)}} \quad (17)$$

или если $(\log n)^{D(n)} < p^*(G)$ и $2^{n-\lambda_\varepsilon} \leq q(G) < \frac{2^n}{(\log n)^{D(n)}}$.

Положим функцию $L_{D,\varepsilon}^2(G, n)$ равной:

- 1) $] \log \chi(G)[$, если $p^*(G) \leq \log n - D(n)$;
- 2) $] \log \max(\chi(G), 2D(n))[$, если $\log n - D(n) < p^*(G) \leq \log n + D(n)$;
- 3) $] \log p^*(G)[$, если $\log n + D(n) < p^*(G) \leq (\ln n) \log \ln n$;

- 4) $[S_1(2 \cdot \binom{p^*(G)}{2}, n) + \varepsilon] + 1$, если $(\ln n) \log \ln n < p^*(G) \leq (\log n)^{D(n)}$;
 5) $[S_1(q(G), n) + \varepsilon] + 1$, если $p^*(G) > (\log n)^{D(n)}$, $q(G) \leq \frac{2^n}{(\log n)^{D(n)}}$.

Пусть x такое, что $1 \leq x \leq 2 \log n$ и $q(G) = \frac{2^n \cdot \ln \binom{n}{x}}{2^x - 1}$. Тогда положим функцию $\hat{L}_{D,\varepsilon}^1(G, n)$ равной

- 1) $[n - x + \varepsilon]$,
 если $\frac{2^n}{(\log n)^{D(n)}} \leq q(G) \leq 2^n \ln n$;
 2) $n - 1$,
 если $2^n \ln n \leq q(G) \leq 2^n (\ln n + D(n))$;
 3) n ,
 если $q(G) \geq 2^n (\ln n + D(n))$.

Положим функцию $\hat{L}_{D,\varepsilon}^2(G, n)$, равной:

- 1) $\hat{L}_{D,\varepsilon}^1(G, n) + 1$,
 если $\frac{2^n}{(\log n)^{D(n)}} \leq q(G) \leq 2^n (\ln n + D(n))$;
 2) $\hat{L}_{D,\varepsilon}^1(G, n)$,
 если $q(G) \geq \frac{2^n}{(\log n)^{D(n)}}$.

Следующее утверждение описывает поведение длин минимальных тестов.

Теорема 5. *Существует $\varepsilon \in (0, \frac{1}{16})$ и $B \in \mathcal{B}$, $D(n) \leq \log \log n$, что*

- 1) *если $1 \leq q(G) \leq \frac{2^n}{(\log n)^{D(n)}}$, то*

$$L_{D,\varepsilon}^1(G, n) \leq_n^{n.б.} L_{D,n}^{\min} \leq_n^{n.б.} L_{D,\varepsilon}^2(G, n); \quad (18)$$

- 2) *если $(\log n)^{D(n)} \leq q(G) \leq \frac{2^n}{(\log n)^{D(n)}}$ и $q(G) \geq p^*(G)^{1+\varepsilon}$, то*

$$[S(q(G), n) + \varepsilon] \leq_n^{n.б.} L_{b,n}^{\min} \leq_n^{n.б.} [S(q(G), n) + \varepsilon] + 1; \quad (19)$$

- 3) *если $q(G) \geq \frac{2^n}{(\log n)^{D(n)}}$, то*

$$\hat{L}_{D,\varepsilon}^1(G, n) \leq_n^{n.б.} \hat{L}_{D,n}^{\min} \leq_n^{n.б.} \hat{L}_{D,\varepsilon}^2(G, n). \quad (20)$$

Опишем процедуру построения тупиковых тестов. Положим $L_{m,n} = L_m \otimes L_n$. Под матрицей T , имеющей m строк и n столбцов, наряду с ранее введенными понятиями, понимаем также отображение $T : L_{m,n} \rightarrow E$. Полагаем, что $\Psi_{m,n} = E^{L_{m,n}}$, $\Psi^* = \bigcup_{(m,n) \in N^2} \Psi_{m,n}$.

Если $T \in \Psi_{m,n}$, то через \hat{T} обозначим матрицу из $\Psi_{L_{m,n}}$ такую, что $(\hat{T}(k))(l) = T(k, l)$ для любой пары (k, l) из $L_{m,n}$. Набор x назовем тестом или тупиковым тестом матрицы T из $\Psi_{m,n}$, если он является тестом или, соответственно, тупиковым тестом матрицы \hat{T} . Полагаем

$$\Lambda_{m,n} = (E^m \setminus \{0^m\}) \otimes (E^n \setminus \{0^n\}). \quad (21)$$

Каждой паре (a, b) из $\Lambda_{m,n}$ соответствует подрешетка $a^{-1}(1) \otimes b^{-1}(1)$ решетки $L_{m,n}$. Если имеется графическое изображение матрицы T из $\Psi_{m,n}$ и пара (a, b) из $\Lambda_{m,n}$, то, удалив строки, соответствующие нулям набора a , и столбцы, соответствующие нулям набора b , получим график таблицы $T \circ (\eta_a, \eta_b)$ из $\Psi_{|a|, |b|}$, где $(T \circ (\eta_a, \eta_b))(i, j) = T(\eta_a(i), \eta_b(j))$. Считаем, что $T^{(a,b)}$, $T^{(a,b)} = T|_{a^{-1}(1) \otimes b^{-1}(1)}$. Набор x из E^n , $x \leq b$ называется тестом или тупиковым тестом подматрицы $T_{(a,b)}$, если набор $x \circ \eta_b$ является тестом или, соответственно, тупиковым тестом матрицы $T^{(a,b)}$.

Под схемой работы алгоритма понимается оператор, ставящий в соответствие каждой матрице набор ее фрагментов, просматриваемых алгоритмом в процессе работы.

Пара $A = (\Sigma, F)$, где Σ каждой паре (m, n) натуральных чисел ставит в соответствие множество $\Sigma(m, n)$, $\Sigma(m, n) \subseteq \Lambda_{m,n}$, а F — функция из E^{Ψ^*} , называется схемой алгоритма, если выполнено:

- 1) $\{1^m\} \otimes (E_2^n \setminus \{0^n\}) \subseteq \Sigma(m, n)$, $(m, n) \in N^2$;
- 2) если $F(T) = 1$, то T — тупиковая тестовая матрица, то есть набор из всех единиц является ее единственным тупиковым тестом;
- 3) если $(a, b) \in \Sigma(m, n)$ и x — тупиковый тест подматрицы $T_{(a,b)}$, то существует c , $c \leq a$, такое, что

$$(c, x) \in \Sigma(m, n); \quad (22)$$

$$F(T^{(c,x)}) = 1. \quad (23)$$

Схема $A = (\Sigma, F)$ называется локальной, если для любой пары (a, b) из $\Sigma(m, n)$ верно: если $c \leq a$, $x \leq b$, то $(c, x) \in \Sigma(m, n)$ точно тогда, когда

$$(c \circ \eta_a, x \circ \eta_b) \in \Sigma(|a|, |b|). \quad (24)$$

Введем граф $G_A(T)$ с множеством вершин

$$V_A(T) = \{(a, b) \mid (a, b) \in \Sigma(m, n), F(T^{(a,b)}) = 1\} \cup \{(0^m, 0^n)\} \quad (25)$$

и множеством ребер $X_A(T)$, состоящим из таких пар вершин $\{(a_1, b_1), (a_2, b_2)\}$, что $a_1 < a_2$, $b_1 < b_2$, $|b_2| = |b_1| + 1$; и, если $a_1 < a < a_2$, $(a, b_1) \in \Sigma(m, n)$, то набор b_1 — тупиковый тест для $T(a, b_1)$.

Схема $A = (\Sigma, F)$ называется связной, если для любой матрицы T из Ψ^* граф $G_A(T)$ — связный.

Введем два функционала сложности

- 1) $\mu(A, T) = |V_A(T)|$,
- 2) $\mu^*(A, T)$ равно числу таких наборов x , что для некоторого c верно, что $(c, x) \in V_A(T)$.

Если A и B — схемы алгоритмов, то пишем $A \leq B$, когда для любой матрицы T из Ψ^* верно

$$\mu(A, T) \leq \mu(B, T). \quad (26)$$

Аналогично определяется и неравенство $A \leq^* B$.

Этим требованиям удовлетворяет, например, схема алгоритма Е. В. Дюковой, указанная в § 1.

Именно, там $A^* = (\Sigma^*, F^*)$, $\Sigma^*(m, n) = \Lambda_{m,n}$, $F^*(T) = 1$ точно тогда, когда T — квадратная тупиковая тестовая матрица. Схема A^* — связная и локальная.

Пусть $f \in \mathcal{H} = \{g \mid g \in N^N, g(n) \leq n\}$. Определим систему наборов $C_f^{k,p} \in E^k$, $p \in L_k$, $k \in N$, соотношениями

$$1^k = C_f^{k,k} > C_f^{k,k-1} > \dots > C_f^{k,1} > C_f^{k,0} = 0^k, \quad (27)$$

$$(C_f^{k,p} - C_f^{k,p-1})(\eta_{C_f^{k,p}}(f(p))) = 1. \quad (28)$$

Для схемы алгоритма $D(f)$ полагаем:

$$\sum_{D(f)} (k, n) = \{(C_f^{k,p}, x) \mid p \in L_k, x \in E^n \setminus \{0^n\}\}; \quad (29)$$

при $T \in \Psi_{k,n}$, $F_{D(f)}(T) = 1$, если T — тупиковая тестовая, а $T^{(C_f^{k,k-1}, 1^n)}$ — не тупиковая матрицы.

Схема $D(f)$ — связанная и локальная.

Теорема 6. Для любой связанной локальной схемы алгоритма A существует f из \mathcal{H} такая, что $D(f) \leq A$, $D(f) \leq^* A$.

Для всех f из \mathcal{H} схема $D(f)$ с точностью до перестановки строк таблицы эквивалентна схеме $D = D(f_0)$, где f_0 — тождественная функция.

Для любой T из Ψ^* граф $G_D(T)$ — дерево.

Опишем алгоритм построения тупиковых тестов, соответствующих схеме D . Если $S \in L_n^{L_l}$, то \hat{S} такой набор из E^n , что $\hat{S}^{-1}(1) = S(L_l)$. Текущими параметрами работы D -алгоритма с матрицей T из $\Psi_{m,n}$ «в момент t », $t \in N$, является пятерка (t, K_t, S_t, l_t, A_t) , где $l_t \in N$, $K_t \in L_m^{L_{l_t}}$, $S_t \in L_n^{L_{l_t}}$, $A_t \in \Psi_{l_t, n}$.

- 1) Если $\hat{T}(1) \in 1^n$, то T не имеет тестов. Конец.
- 2) Полагаем $t = 1$, $l_t = 1$, $K_t(1) = 1$, $\hat{A}_t(1) = \hat{T}(1) \oplus 1^n$, $S_t(1) = \min(\hat{T}(1))^{-1}(0)$. Перейти к п. 3).
- 3) Полагаем $\varepsilon = (\hat{T}^{(1^m, \tilde{S}_t)})^{-1}(1^{l_t})$. Если $\varepsilon = \emptyset$, \tilde{S}_t — тупиковый тест, переходим к п. 5); если $\varepsilon \neq \emptyset$, то перейти к п. 4);
- 4) Полагаем $k = \min \varepsilon$; $\varepsilon_i = (\hat{T}^{(C_{f_0}^{m,k}, \tilde{S}_t)})^{-1}(1_i^{l_t})$;

$$y = \left(\bigotimes_{i=1}^{l_t} \left(\bigvee_{r \in \varepsilon_i} \hat{T}(r) \right) \right) \& (\hat{T}(k) \oplus 1^n). \quad (30)$$

Если $y = 0^n$, то перейти к п. 5). В противном случае увеличиваем t на 1 и получаем:

$$l_t = l_{t-1} + 1, \quad \hat{A}_t(l_t) = y, \quad k_t(l_t) = k; \quad (31)$$

$$(K_t, S_t, \hat{A}_t) \Big|_{L_{l_{t-1}}} = (K_{t-1}, S_{t-1}, \hat{A}_{t-1}); \quad (32)$$

$$S_t(l_t) = \min y^{-1}(1); \quad (33)$$

и переходим к п. 3).

- 5) Полагаем $C_r = (\hat{A}_t(r))^{-1} \cap (L_n \setminus L_{S_t(r)})$.
 Если для всех r из L_t верно $C_r = \emptyset$, то конец. В противном случае полагаем $j_0 = \min C_{r_0}$, где $r_0 = \min\{r | C_r \neq \emptyset\}$; увеличиваем t на 1 и полагаем:

$$S_t(l_t) = \min y^{-1}(1); \quad (34)$$

$$S_t|_{L_{r_0-1}} = S_{t-1}|_{L_{r_0-1}}; \quad (35)$$

$$S_t(r_0) = j_0, \quad l_t = r_0; \quad (36)$$

$$(K_t, \hat{A}_t) = (K_{t-1}, \hat{A}_{t-1})|_{L_{r_0}}; \quad (37)$$

переходим к п. 3).

Если π — нумерация ребер графа G (биекция $X(G)$ на $L_{q(G)}$), то под работой пары (A, π) , где A — схема алгоритма, на паре (T, G) из $\Psi_{b,n}$ понимаем работу схемы A на таблице $T_G \circ \pi^{-1}$. Полагаем $\mu((A, \pi), (T, G)) = \mu(A, T_G \circ \pi^{-1})$, и аналогично для μ^* . Здесь матрицы T и \hat{T} отождествляются.

Теорема 7. *Если каждому графу G соответствует нумерация его ребер π_G и $\log q(G) \ll_n \log^2 n$, то*

$$\mu((D, \pi_G), (T, G)) = \mu^*((D, \pi_G), (T, G)) \lesssim_n^{n.в.} \varphi_{G,n}^{TT}(T, G). \quad (38)$$

3. Короткие тесты

В этом параграфе излагаются результаты А. А. Кибкало по задачам нахождения информационных весов признаков и процедур эффективного распознавания, а также быстрого построения основных семейств тестов.

В отличие от А. Е. Андреева здесь рассматривается случай, когда матрица T' состоит только из двух подматриц T'_1 и T'_2 , хотя в принципе получаемые результаты распространяемы и на общую ситуацию.

В качестве главных семейств тестов для T , называемых опорными, выбираются:

- 1) множество всех тестов,
- 2) множество всех тупиковых тестов,

- 3) множество тестов длины не более r ,
- 4) множество тупиковых тестов длины не более r .

Для этих опорных множеств найдена асимптотика веса признака для почти всех таблиц с заданной долей различаемых этим признаком пар объектов, не лежащих в одной подтаблице T'_i . В частности, установлено, что вес признака по всем тестам не зависит от доли различаемых им пар и почти всегда асимптотически равен $\frac{1}{2}$.

Вес по всем тупиковым тестам убывает с ростом доли пар строк, различаемых признаком. Если же r таково, что длину не более r имеет лишь асимптотически небольшое число тестов, и в типичной ситуации почти все такие тесты являются тупиковыми, то веса признака по тестам длины не более r и тупиковым тестам длины не более r асимптотически совпадают, и, в отличие от веса по всем тупиковым тестам, растут с ростом доли пар строк из разных подматриц, различаемых признаком. Такие тесты считаются короткими.

Пусть Ω — оператор, сопоставляющий таблице T' опорное множество тестов $\Omega(T')$. Оператор Ω , а также алгоритм распознавания, основанный на $\Omega(T')$, считается устойчивым, если отношение

$$\frac{|\Omega(T') \cap \Omega(T^*)|}{|\Omega(T') \cup \Omega(T^*)|} \quad (39)$$

близко к единице при небольших отличиях таблиц T и T^* . Для оператора Ω типа а) почти все наборы признаков являются тестами и, как отмечалось выше, он тем самым не представляет интереса.

Для типов б), в) и г) установлено следующее. Оказалось, что при малых искажениях обучающей таблицы T' множество всех тупиковых тестов почти полностью меняется, в то время как множество коротких тестов почти полностью сохраняется. Кроме того, множество «очень коротких» тестов, длина которых близка к минимальной, также изменяется почти полностью.

Указанные свойства весов признаков и устойчивости опорного множества тестов означают, что основную информацию о различиях между подматрицами T'_1 и T'_2 несет множество коротких тестов, а тесты большой длины и близкие к минимальным представляют собой случайный фон.

А. А. Кибкало построены алгоритмы D_1 и D_2 отыскания коротких тестов, являющихся асимптотически оптимальными при существенно более общих предположениях, чем все остальные.

Алгоритм D_1 отыскания всех коротких тестов является модификацией алгоритма А. Е. Андреева, а алгоритм D_2 предназначен для построения всех тестов длины не более, чем r .

Пусть $\Psi_{v_1, v_2, n}$ — семейство бинарных матриц $T = (T_1, T_2)$, таких что в подматрице T_i содержится v_i строк, $i = 1, 2, \dots$ и число столбцов у T_1 и T_2 равно n . Обозначим через \hat{T} таблицу сравнения для T .

Пусть $\varphi_{v_1, v_2, n, \tilde{x}}^T$, где $\tilde{x} \in E_2^n \setminus \{\tilde{\sigma}^n\}$ — функция, определенная на множестве $\Psi_{v_1, v_2, n}$ и принимающая значение 1, если \tilde{x} является тестом для T , и значение 0 в противном случае.

Для r через $\varphi_{v_1, v_2, n, \tilde{x}}^T$ обозначим функцию, равную числу тестов длины r у T из $\Psi_{v_1, v_2, n}$, а через $\varphi_{v_1, v_2, n}^T$ — функцию, равную числу всех тестов у T , тогда если $E_2^{n, r}$ — множество всех наборов из E_2^n содержащих ровно r единиц, то

$$\varphi_{v_1, v_2, n, r}^T = \sum_{\tilde{x} \in E_2^n} \varphi_{v_1, v_2, n, \tilde{x}}^T, \quad \varphi_{v_1, v_2, n}^T = \sum_{i=1}^n \varphi_{v_1, v_2, n, r}^T. \quad (40)$$

Аналогично определяется функция $\varphi_{v_1, v_2, n, \tilde{x}}^{TT}$, которая равна 1, если \tilde{x} — тупиковый тест для T , и 0 в остальных случаях. Определим функции $\varphi_{v_1, v_2, n, r}^{TT}$, равной числу тупиковых тестов длины r , $\varphi_{v_1, v_2, n}^{TT}$ — числу всех тупиковых тестов, $\varphi_{v_1, v_2, n, r}^{KT}$ — числу тестов длины не более r , φ_{v_1, v_2}^{KTT} — числу всех тупиковых тестов длины не более r .

Для этих функций выполнено

$$\varphi_{v_1, v_2, r}^{TT} = \sum_{\tilde{x} \in E_2^{n, r}} \varphi_{v_1, v_2, \tilde{x}}^{TT}, \quad \varphi_{v_1, v_2, r}^{TT} = \sum_{r=1}^n \varphi_{v_1, v_2, n, r}^{TT}, \quad (41)$$

$$\varphi_{v_1, v_2, n, r}^{KT} = \sum_{j=1}^r \varphi_{v_1, v_2, n}^T, \quad \varphi_{v_1, v_2, n, r}^{KTT} = \sum_{j=1}^r \varphi_{v_1, v_2, n, j}^{TT}. \quad (42)$$

Если $\nabla \in \{T, TT\}$, $\Delta \in \{T, TT, KT, KTT\}$, $i \in L_n$, $r \in L_n$, то полагаем

$$\varphi_{v_1, v_2, n, r}^{\nabla, i} = \sum_{\tilde{x} \in E_2^{n, r}, \tilde{x}(i)=1} \varphi_{v_1, v_2, n, \tilde{x}}^{\nabla} \quad (43)$$

$$\varphi_{v_1, v_2, n}^{\nabla, i} = \sum_{r=1}^n \varphi_{v_1, v_2, r}^{\nabla, i}, \quad (44)$$

$$\varphi_{v_1, v_2, n, r}^{K, \nabla, i} = \sum_{j=1}^r \varphi_{v_1, v_2, n, j}^{\nabla, i}. \quad (45)$$

Определим на $\Psi_{v_1, v_2, n}$ функции $\psi_{v_1, v_2, n, r}^{\Delta, i}$, $\psi_{v_1, v_2, n}^{\nabla, i}$, равные доле тестов соответствующего вида, содержащих i -й признак,

$$\psi_{v_1, v_2, r}^{\Delta, i} = \frac{\varphi_{v_1, v_2, n, r}^{\Delta, i}}{\varphi_{v_1, v_2, n, r}^{\Delta}}, \quad (46)$$

$$\psi_{v_1, v_2, n}^{\nabla, i} = \frac{\varphi_{v_1, v_2, n}^{\nabla, i}}{\varphi_{v_1, v_2, n}^{\nabla}}. \quad (47)$$

Функции $\psi_{v_1, v_2, n}^{\nabla, i}$ и $\psi_{v_1, v_2, n, r}^{\Delta, i}$ называются весами i -го признака в T из $\Psi_{v_1, v_2, n}$, а именно:

- $\psi_{v_1, v_2, n}^{T, i}$ — вес по всем признакам;
- $\psi_{v_1, v_2, n}^{TT}$ — вес по всем тупиковым тестам;
- $\psi_{v_1, v_2, n, r}^{T, i}$ — вес по всем тестам длины r ;
- $\psi_{v_1, v_2, n, r}^{TT, i}$ — вес по всем тестам длины не более r ;
- $\psi_{v_1, v_2, n, r}^{KTT, i}$ — вес по всем тупиковым тестам длины не более r .

Можно считать, что $i = 1$ и индекс i далее можно опускать.

Пусть $\tilde{x} \in E_2^n \setminus \{\tilde{o}^n\}$ и $|\tilde{x}|$ равно числу единиц в наборе \tilde{x} . Пусть $\eta_{\tilde{x}} : L_{|\tilde{x}|} \rightarrow L_n$, полагая $\eta_{\tilde{x}}(i) = k_i$, $i \in L_{|\tilde{x}|}$, где k_i — номер i -й единицы набора \tilde{x} .

Если в наборе $\tilde{y} \in E_2^n$ вычеркнуть координаты, соответствующие нулям набора \tilde{x} , то полученный набор обозначим через $\tilde{y} \circ \eta_{\tilde{x}}$.

Для \tilde{x} из E_2^n определим отображение $\pi_{\tilde{x}} : E_2^n \rightarrow E_2^{|\tilde{x}|}$ так, что $\pi_{\tilde{x}}(\tilde{y}) = \tilde{y} \circ \eta_{\tilde{x}}$.

Если первые $l, l \in L_n$, координат набора \tilde{x} из E_2^n равны 1, а остальные равны 0, то отображение $\pi_{\tilde{x}}$ записываем $\pi_{n,l}$.

При $\alpha \in (0, 1)$, $\varepsilon \in (0, \frac{1}{2})$ обозначим через $\Psi_{v_1, v_2, n}^{\alpha, \varepsilon}$ множество матриц T из $\Psi_{v_1, v_2, n}$ таких, что

$$\varepsilon m_i \leq |(\pi_{n,1} \circ T_i)^{-1}(1)| \leq (1 - \varepsilon) m_i, \quad i \in L_2, \quad (48)$$

$$|(\pi_{n,1} \circ T_i)^{-1}(1)| = \alpha m_1 m_2. \quad (49)$$

Рассматриваются такие α , что $\Psi_{v_1, v_2, n}^{\alpha, \varepsilon}$ непусто. Ясно, что для каждого ε из $(0, \frac{1}{2})$ такие α лежат на отрезке $[2\varepsilon - 2\varepsilon^2, 1 - 2\varepsilon + 2\varepsilon^2]$, который обозначаем I^ε .

Через $\psi_{v_1, v_2, n}^{\alpha, i}$, $\psi_{v_1, v_2, n}^{\alpha, TT}$, $\psi_{v_1, v_2, n, r}^{\alpha, T}$, $\psi_{v_1, v_2, n, r}^{\alpha, TT}$, $\psi_{v_1, v_2, n, r}^{\alpha, KT}$, $\psi_{v_1, v_2, n, r}^{\alpha, KTT}$ обозначим сужение соответствующих весовых функций на множество $\Psi_{v_1, v_2, n}^{\alpha, \varepsilon}$.

Пусть

$$r_{1,k}^\varepsilon(m) =] \log m - (1 - \varepsilon) \log \log m[, \quad (50)$$

$$r_{2,k}^\varepsilon(m) = [\log m - (1 + \varepsilon) \log \log \log m], \quad (51)$$

$$\tilde{r}_{1,k}^\varepsilon(m) = [\log m - (1 + \varepsilon) \log \log m]. \quad (52)$$

Теорема 8. Пусть c_1, c_2, ε — константы, $0 < c_1 < 1 < c_2$, $\varepsilon \in (0, \min(\frac{1}{32}, \frac{c_1}{20c_2}))$, тогда если $n^{c_1} \leq m_1 m_2 \leq n^{c_2}$, $m_1 \asymp_n m_2$, $\alpha_1, \alpha_2 \in I^\varepsilon$, то

1) существует такая функция $\psi^\alpha(n, m_1 m_2)$, что

$$\psi_{v_1, v_2, n}^{\alpha, TT} \sim \psi^\alpha(n, m_1 m_2), \quad (53)$$

а при $\alpha_1 < \alpha_2$ выполнено

$$\psi^{\alpha_1}(n, m_1 m_2) \ll_n \psi^{\alpha_2}(n, m_1 m_2); \quad (54)$$

2) $\psi_{v_1, v_2, n}^{\alpha, T} \sim_n^{н.б.} \frac{1}{2}$;

3) существует такая функция $\psi_k^\alpha(n, m_1 m_2, r)$, что при

$$r \in [\tilde{r}_{1,k}^\varepsilon(m_1 m_2), r_{2,k}^\varepsilon(m_1 m_2)] \quad (55)$$

выполнено

$$\begin{aligned} \psi_{v_1, v_2, n, r}^{\alpha, KT} \sim_n^{n.б.} \psi_{v_1, v_2, n, r}^{\alpha, KTT} \sim_n^{n.б.} \psi_{v_1, v_2, n, r}^{\alpha, T} \sim_n^{n.б.} \\ \sim_n^{n.б.} \psi_{v_1, v_2, n, r}^{\alpha, TT} \sim_n^{n.б.} \psi_k^\alpha(n, m_1 m_2, r); \end{aligned} \quad (56)$$

если $\alpha_1 < \alpha_2$ и $r \in [\tilde{r}_{1,k}^\varepsilon(m_1 m_2), r_{2,k}^\varepsilon(m_1 m_2)]$, то

$$\psi_k^{\alpha_1}(n, m_1 m_2, r) \gg_n \psi_k^{\alpha_2}(n, m_1 m_2, r); \quad (57)$$

если $\alpha_1 < \frac{1}{2}$ и $r \in [\tilde{r}_{1,k}^\varepsilon(m_1 m_2), \tilde{r}_{2,k}^\varepsilon(m_1 m_2)]$, то

$$\psi_k^\alpha(n, m_1 m_2, r) \sim_n 1. \quad (58)$$

Таким образом Теорема 8 описывает асимптотическое поведение весов признаков по различным множествам тестов в зависимости от доли различаемых признаком пар объектов, не лежащих в одном классе. Из нее вытекает следующее:

- 1) Вес признака по всем тупиковым тестам имеет тенденцию к уменьшению с увеличением доли различаемых признаком пар объектов.
- 2) Вес признака по всем тестам не зависит от доли различаемых признаком пар объектов и почти всегда равен $\frac{1}{2}$.
- 3) Вес признака по всем коротким тестам (коротким тупиковым тестам) растет с увеличением доли различаемых признаком пар объектов.
- 4) Вес признака, различающего больше половины пар объектов по множеству тестов, длина которых близка к минимальной, почти всегда равен единице.

Поэтому в качестве опорного множества тестового алгоритма распознавания целесообразно использовать множество всех коротких тестов, то есть тестов, длина которых не превосходит некоторого числа из интервала $(\log m - \log \log m, \log m - \log \log \log m)$, а в качестве меры информативности признака — его вес по множеству коротких тестов.

Рассмотрим вопрос устойчивости опорных множеств тестов различного вида при малых искажениях обучающей таблицы.

С этой целью определяется множество $\Psi_{v_1, v_2, n}^2 = \Psi_{v_1, v_2, n} \otimes \Psi_{v_1, v_2, n}$ и рассматривается конечное вероятностное пространство со следующим распределением

$$p\{(T, T^*)\} = 2 \cdot p^{\tilde{\rho}(T, T^*)} (1-p)^{(|V_1|+|V_2|)n - \tilde{\rho}(T, T^*) - (|V_1|+|V_2|)n}, \quad (59)$$

где $\rho(T, T^*)$ — число отличий T и T^* из $\Psi_{v_1, v_2, n}$, то есть

$$\tilde{\rho}(T, T^*) = \sum_{i=1}^r \sum_{a \in V_i} \rho(T_i(a), T_j^*(a)). \quad (60)$$

Это распределение моделирует ситуацию, когда каждая компонента матрицы T независимо от других с вероятностью p меняет свое значение на противоположное, и в результате изменений возникает таблица T^* .

Определим на $\Psi_{v_1, v_2, n}^2$ функции $\psi_{v_1, v_2, n}^{TT, \&}$, $\psi_{v_1, v_2, n}^{TT, \vee}$, $\psi_{v_1, v_2, n, r}^{TT, \&}$, $\psi_{v_1, v_2, n, r}^{TT, \vee}$, $\psi_{v_1, v_2, n, r}^{KTT, \&}$, $\psi_{v_1, v_2, n, r}^{KTT, \vee}$, $\psi_{v_1, v_2, n, r}^{T, \&}$, $\psi_{v_1, v_2, n, r}^{T, \vee}$, $\psi_{v_1, v_2, n, r}^{KT, \&}$, $\psi_{v_1, v_2, n, r}^{KT, \vee}$ так, что функции со значком $\&$ задают число наборов из E_n^2 , являющихся тестами соответствующего вида как для правильно заданной, так и для искаженной матрицы, а функции со значком \vee — число наборов, являющихся тестами хотя бы одной из этих таблиц.

Множество тестов некоторого из указанных видов называется устойчивым к искажениям, если $\varphi^{*, \&} \sim_n^{п.в.} \varphi^{*, \vee}$, где φ^* — функция, задающая число тестов соответствующего вида.

Теорема 9. Если $\varepsilon \in (0, \min(\frac{1}{32}, \frac{c_1}{20c_2}))$, $0 < c_1 < 1 < c_2$, $n^{c_1} \leq m_1 m_2 \leq n^{c_2}$, $(\log m_1 m_2)^{-2} \ll p \ll (\log m_1 m_2)^{-(1+\varepsilon)}$, при $n \rightarrow \infty$, $\nabla \in \{T, TT, KT, KTT\}$, то

- 1) при $r =] \log m_1 m_2 - \varepsilon \log \log m_1 m_2 [$ или $r =] \log m_1 m_2 - b \log \log m_1 m_2 [$, $b > 1 + \varepsilon$ выполнено почти всегда при $n \rightarrow \infty$

$$\psi_{v_1, v_2, n, r}^{\nabla, \&} \sim_n^{н.с.} \psi_{v_1, v_2, n, r}^{\nabla, \vee} \sim_n^{н.с.} (C_n^r) e^{(-\frac{m_1 m_2}{2r})}; \quad (61)$$

- 2) $\psi_{v_1, v_2, n}^{TT, \&} \ll_n^{н.с.} \psi_{v_1, v_2, n}^{TT, \vee}$;

- 3) при $r =] \log m_1 m_2 - a \log \log m_1 m_2 [$, $a \in (1 + \varepsilon, 2 - \varepsilon)$ почти всегда при $n \rightarrow \infty$ выполнено

$$\psi_{v_1, v_2, n, r}^{\nabla, \&} \ll_n^{н.с.} \psi_{v_1, v_2, n, r}^{\nabla, \vee}. \quad (62)$$

Эта теорема описывает асимптотическое поведение множества тупиковых тестов матриц из $\Psi_{v_1, v_2, n}$ при искажениях таблиц, задаваемых распределением вероятностей p . При таких искажениях почти все тупиковые тесты правильно заданной матрицы T не являются тупиковыми тестами искаженной таблицы, и, наоборот, в то время как множество коротких тестов практически не изменяется. Кроме того, множество тестов (тупиковых тестов), длины которых близки к минимальной, тоже изменяется почти полностью. Таким образом, распознающие алгоритмы, использующие в качестве опорного множества множество всех коротких тестов матриц, будут устойчивыми к малым искажениям обучающей информации.

А. А. Кибкало предложены алгоритмы построения коротких тестов.

Под нумерацией строк таблицы сравнения понимается взаимно-однозначное отображение

$$\pi : x(G_{v_1, v_2}) \rightarrow L_{m_1 m_2}. \quad (63)$$

Пусть $\hat{T}_\pi = \hat{T} \circ \pi^{-1}$ и $\Psi_{m, n}^*$ — множество матриц S , $S : L_m \rightarrow E^n$, положим $\Psi^* = \bigcup_{(m, n) \in N^2} \Psi_{m, n}^*$. Пусть $S \in \Psi^*$ и A — алгоритм построения тестов (тупиковых тестов). Под сложностью работы алгоритма $\mu^T(A, S)(\mu^{TT}(A, S))$ понимается, как и ранее при описании результатов А. Е. Андреева, число наборов из E_2^n , проверяемых в алгоритме на принадлежность к множеству тестов (тупиковых тестов) матрицы S .

Предлагаются два алгоритма построения коротких тестов.

Алгоритм $D_1(r)$ строит все тупиковые тесты длины не более r и является модификацией алгоритма А. Е. Андреева.

Алгоритм $D_2(r)$ строит все тупиковые тесты длины не более r .

Набор S отождествим с отображением $S : L_l \rightarrow L_n$. Пусть $\hat{s} \in E^n$, $\hat{s}^{-1}(1) = S(L_l)$ и $\hat{T}_\pi^{\hat{s}} = \hat{T}_\pi \circ \pi_{\hat{s}}$. Доопределим матрицу \hat{T}_π на L_{m+1} так, что $\hat{T}_\pi(m+1) = \delta^n$. Текущими параметрами алгоритма $D_2(r)$ являются $t, k_t, S_t, l_t, S_t, M_t$, где t — номер проверяемого алгоритмом набора из E^n , l_t — длина этого набора.

На шаге номер t алгоритм проверяет, будет ли набор \tilde{S}_t из E^n , где $S_t : L_{l_t} \rightarrow L_n$, тестом матрицы \hat{T}_π . Набор $K_t, K_t : L_{l_t} \rightarrow L_{m+1}$ и матрицы A_t, M_t из $\Psi_{l_t, n}^*$ служат для организации работы алгоритма.

Работа алгоритма $D_2(r)$ на матрице \hat{T}_π , $r > 1$, осуществляется следующим образом:

- 1) Если $\hat{T}_\pi(1) = \tilde{1}^n$, то \hat{T}_π не имеет тестов, перейти к п. 7). В противном случае перейти к п. 2).
- 2) Полагаем $t = 1$, $l_t = 1$, $K_t(1) = 1$, $A_t(1) = \hat{T}_\pi(1) \oplus \tilde{1}^n$, $M_t(1) = \tilde{\delta}^n$, $S_t(1) = \min A_t(1)^{-1}(0)$. Перейти к п. 3).
- 3) Полагаем ε равными $(\hat{T}_\pi^{\hat{s}_t})^{-1}(\tilde{1}^{l_t})$, если $K_t(l_t) < m + 1$ и $m + 1$ в противном случае.
Если $\varepsilon \cap L_m = \emptyset$, то \hat{s}_t — тест.
Если $l_t = r - 1$, то перейти к п. 5).
В противном случае перейти к п. 4).
- 4) Положим $k = \min \varepsilon$ и $\tilde{y} = (\hat{T}_\pi(k) \vee M_t(l_t)) \oplus \tilde{1}^n$. Если $\tilde{y} = \tilde{\delta}^n$, то перейти к п. 6).
В противном случае полагаем $t = t + 1$, $l_t = l_{t-1} + 1$, $A_t(l_t) = \tilde{y}$, $S_t(l_t) = \min \tilde{y}^{-1}(1)$, $K_t(l_t) = K$, $M_t(l_t) = M_{t-1}(l_t - 1)$, $(K_t, S_t, A_t, M_t)|_{L_{l_t-1}} = (K_{t-1}, S_{t-1}, A_{t-1}, M_{t-1})$, перейти к п. 3).
- 5) Полагаем $\tilde{y} = ((\bigvee_{i \in \varepsilon} \hat{T}(i)) \vee M_t(l_t)) \oplus \tilde{1}^n$.
Для любого $i \in \tilde{y}^{-1}(1)$ объект $\hat{s}_t(+)\tilde{\delta}_i^n$ — тест.
Полагаем $t = t + |\tilde{y}|$, $(l_t, S_t, A_t, M_t) = (l_{t-|\tilde{y}|}, S_{t-|\tilde{y}|}, K_{t-|\tilde{y}|}, M_{t-|\tilde{y}|})$.
Перейти к п. 6).
- 6) Полагаем для $i \in L_{l_t}$

$$c_j = A_t(j)^{-1}(1) \cap (L_n \setminus L_{S_t(j)}). \quad (64)$$

Если $c_j = \emptyset$ для всех j на L_{l_t} , то перейти к п. 7). В противном случае полагаем $t = t + 1$, $l_t = \max\{j | c_j \neq \emptyset\}$, $M_t(l_t) = M_{t-1}(l_t) \vee \tilde{\delta}_{S_{t-1}(l_t)}^n$, $S_t(l_t) = \min c_{l_t}$, $(S_t, M_t)|_{L_{l_t-1}} = (S_{t-1}, M_{t-1})|_{L_{l_t-1}}$,

$$(K_t, A_t) = (K_{t-1}, A_{t-1})|_{L_{l_t}}. \quad (65)$$

Перейти к п. 3).

- 7) Закончить работу.

Теорема 10. Если π из P_{v_1, v_2} — произвольная нумерация ребер графа G_{v_1, v_2} , T — пара матриц из $\Psi_{v_1, v_2, n}$, то имеет место:

1) при $c > 1$, $\log m \leq (\log m)^c$, $a \in (0, \frac{1}{c})$ и

$$r = \lceil \log m - a \log \log m \rceil \quad (66)$$

$$\mu^{TT}(D_1(r), \hat{T}_\pi) \sim_n^{n.с.} \mu^T(D_2(r), \hat{T}_\pi) \sim_n^{n.с.} \psi_{v_1, v_2, n, r}^{KT}; \quad (67)$$

2) при $\varepsilon \in (0, 1)$, $\log m \leq 2^{(\log n)^{1-\varepsilon}}$, $b \in (1, \frac{1}{1-\varepsilon})$,

$$r = \lceil \log m - b \log \log \log m \rceil \quad (68)$$

выполнено

$$\mu^{TT}(D_1(r), \hat{T}_\pi) \sim_n^{n.с.} \mu^T(D_2(r), \hat{T}_\pi) \sim_n^{n.с.} \psi_{v_1, v_2, n, r}^{KTT} \sim_n^{n.с.} \psi_{v_1, v_2, n, r}^{KT}. \quad (69)$$

4. Тесты для функций k -значной логики

Вычислительные системы строятся из элементов, в которых соответствие между входными и выходными сигналами описывается некоторой логической функцией.

При сборке вычислительных систем и их эксплуатации могут возникать ошибки типа неверных соединений элементов и др. В этом случае вся система начинает работать не так, как проектировалось. Возникает задача проверки того, верно ли функционирует система. В общем случае нет другого пути, кроме, как проверить все соответствия выходных значений системы ее входным воздействиям, что конечно, весьма громоздко и требует больших усилий.

Реально же ошибок в системе происходит не очень много и тем самым установление верности функционирования системы не обязательно связано с большим перебором возможностей.

Основным видом внешнего воздействия для проверки правильности работы системы является подача входных сигналов на систему и проверка выходных реакций на нее.

Иногда удается за счет анализа небольшого числа таких соответствий установить, верно ли работает система. Такие соответствия называют тестовыми. Таким образом, задача проверки правильности работы системы отождествляется с нахождением небольшого числа

тестовых соответствий, по которым и определяется исправность системы.

Здесь мы изложим результаты Г. Р. Погосяна по этому направлению.

Формализуем сказанное.

Отождествляем элемент с функцией $f : E_k^n \rightarrow E_k$, которую называем функцией k -значной логики. Такие функции обычно записываются в виде $f(x_1, x_2, \dots, x_n)$, где переменные x_i , как и сама функция, принимают значения из E_k .

Пусть $x^n = \{x_1, x_2, \dots, x_n\}$ и $N_n = \{1, 2, \dots, n\}$.

Класс всех таких функций обозначают через P_k^n , полагая, что $P_k = \bigcup_{n=1}^{\infty} P_k^n$.

Пусть $\varphi : E_k^n \rightarrow E_k^n$. Это отображение называем φ -ошибкой.

Пусть Φ_k^n — класс всех φ -ошибок от n переменных и $\Phi_k = \bigcup_{n=1}^{\infty} \Phi_k^n$.

С каждой φ -ошибкой и f связываем функцию $f_\varphi(x_1, x_2, \dots, x_n) = f(\varphi(x_1, x_2, \dots, x_n))$, которую называем φ -искажением функции f .

Пусть $\Phi \subseteq \Phi_k^n$ и $\Phi = \{\varphi_1, \varphi_2, \dots, \varphi_s\}$.

Для нас важно установление того, когда $f \neq f_{\varphi_i}$ при любых $i = 1, 2, \dots, s$. Это осуществляется с помощью понятия теста, введенного ранее.

Подмножество $T \subseteq E_k^n$ называем проверяющим тестом для функции $f \in P_k^n$ по отношению к классу ошибок Φ , если для любой ошибки $\varphi \in \Phi$ выполнено либо $f(\alpha) = f(\varphi(\alpha))$ для всех наборов α из E_k^n , либо для некоторого набора β из T выполнено $f(\beta) = f(\varphi(\beta))$.

Пусть $M(f, \Phi)$ — множество всех проверяющих тестов для f относительно Φ и

$$L(f, \Phi) = \min_{T \in M(f, \Phi)} |T|. \quad (70)$$

Тест $T \in M(f, \Phi)$ называется минимальным, если $|T| = L(f, \Phi)$.

Вводим функцию Шеннона для длины проверяющего теста, полагая

$$L(n, \Phi) = \max_{f \in P_k^n} L(f, \Phi). \quad (71)$$

Введем класс неисправностей F_c^k , называемый классом констант, следующим образом

$$F_c^k = \{\varphi_\gamma : \gamma \in E_{k+1}^n \setminus (k, k, \dots, k), \\ \forall \alpha \in E_k^n (\varphi_\gamma(\alpha) = (c_1 * a_1, \dots, c_n * a_n))\}, \quad (72)$$

где $\gamma = (c_1, \dots, c_n)$, $\alpha = (\alpha_1, \dots, \alpha_n)$, $c_i * a_i = c_i$, если $c_i \in E_k^n$ и $c_i * a_i = a_i$, если $c_i = k$.

Теорема 11. *Для любых $n \leq 1$ и $k \leq 2$ имеет место*

$$L(n, F_c^k) = \begin{cases} 2n - 2t, & \text{если } k^{t-1} + t < n \leq k^t + t; \\ (2n - 2t) + 1, & \text{если } n = k^{t-1} + t. \end{cases} \quad (73)$$

Кратностью ошибки $\varphi_\gamma \in F_c^k$ называем число координат набора $\gamma = (c_1, \dots, c_n)$, отличных от k . Множество всех ошибок из F_c^k , имеющих кратность не более, чем p , обозначим через $F_c^k(p)$.

Пусть $M \subseteq E_k$. Положим

$$F_M^k = \{\varphi_\gamma \in F_c^k : \forall i \in N_n (c_i \in M \cup \{k\})\}, \quad (74)$$

где $\gamma = (c_1, \dots, c_n)$. Если $|M| = 1$ и $M = \{\mu\}$, то вместо F_M^k пишем F_μ^k .

Теорема 12. *Для любых $n \geq 1$, $k \geq 2$, $p \in N_n$ имеет место*

$$L(n, F_c^k(p)) = L(n, F_c^k). \quad (75)$$

Теорема 13. *Для любых $n \geq 1$, $k \geq 2$, $\mu \in N_n$ имеет место*

$$L(n, F_\mu^k) = n. \quad (76)$$

Теорема 14. *Для любых $n \geq 1$, $k \geq 2$ и любого подмножества $M \subseteq E_k$ такого, что $|M| \geq 2$, имеет место*

$$L(n, F_M^k) = L(n, F_c^k). \quad (77)$$

Введем класс ошибок F_{st}^k . С каждой ошибкой $\varphi \in F_{st}^k$ однозначно связываем разбиение $S(\varphi)$ множества переменных x_n на непересекающиеся непустые подмножества $Z_1(\varphi), \dots, Z_{q_\varphi}(\varphi)$, где $q_\varphi \in N_{n-1}$. При этом для любого набора α из E_k^n полагаем $\varphi(\alpha) = \beta = (b_1, \dots, b_n)$, где для каждого i из N_n выполнено $b_i = \max\{a_j : x_j \in Z_l(\varphi)\}$, если $x_j \in Z_l(\varphi)$, $l \in N_{q_\varphi}$.

Теорема 15. *Для любых $n \geq 2$, $k \geq 2$, $\mu \in N_n$ имеет место*

$$L(n, F_{st}^k) = n - 1. \quad (78)$$

Кратностью ошибки $\varphi \in F_{st}^k$ называем величину $p(\varphi) = n - q_\varphi$, где q_φ — подмножество множества x_n в разбиении $S(\varphi)$. Пусть $F_{st}^k(p)$ — множество всех ошибок из F_{st}^k , имеющих кратность не более, чем p .

Теорема 16. *Для любых $n \geq 2$, $k \geq 2$ и $p \in N_{n-1}$ имеет место*

$$L(n, F_{st}^k(p)) = L(n, F_{st}^k). \quad (79)$$

В случае, когда $k = 2$, получаем

$$J_n = \{\varphi_\sigma : \sigma \in E_2^n, \forall \alpha \in E_2^n (\varphi_\sigma(\alpha) = \alpha + \sigma)\}, \quad (80)$$

где $\alpha + \sigma$ обозначает набор, получающийся из α и σ их покомпонентным сложением по модулю 2. Помимо этого сложения введем покомпонентное умножение набора α на число a из E_2 .

Множество J_n с операциями $+$, \cdot образует линейное пространство над полем вычетов по модулю 2.

Пусть \tilde{o} — набор из E_2^n , каждая координата которого равна нулю.

Класс инверсий F_{in}^2 определяем так:

$$F_{in}^2 = J_n \setminus \{\varphi_{\tilde{o}}\}. \quad (81)$$

Теорема 17. *Для любого $n \geq 1$ имеет место*

$$2 \cdot \left\lfloor \frac{n-1}{2} \right\rfloor + 1 \leq L(n, F_{in}^2) \leq n. \quad (82)$$

Пусть δ — такой набор из E_2^n , у которого i -я координата равна единице, а все остальные равны нулю. Пусть

$$F_{in}^2(1) = \{\varphi_{\delta^1}, \dots, \varphi_{\delta^n}\}. \quad (83)$$

Теорема 18. *Для любого $n \geq 1$ имеет место*

$$L(n, F_{in}^2(1)) = n - t, \quad (84)$$

где t определяется из соотношения

$$2^{t-1} + t \leq n \leq 2^t + t. \quad (85)$$

Интерес представляет рассмотрение классов разнотипных ошибок, в отдельности уже исследованных выше.

Оказывается, что для любой функции $f \in P_2^n$ и любого теста $T \in M(f, F_c^2)$ имеет место $T \in M(f, F_{in}^2(1))$, отсюда вытекает, что

$$L(n, F_c^2 \cup F_{in}^2(1)) = L(n, F_c^2). \quad (86)$$

Функция $f \in P_k^n$ называется инвариантной относительно некоторой ошибки φ , если для любого $\alpha \in E_k^n$ имеет место $f(\alpha) = f(\varphi(\alpha))$, в противном случае f называется чувствительной к ошибке φ .

Функция f инвариантна (чувствительна) относительно множества ошибок Φ , если она инвариантна (чувствительна) относительно каждой ошибки из Φ .

Показано, что для каждого Φ из $\{F_c^k, F_{st}^k, F_{in}^2\}$ почти все функции из P_k^n являются чувствительными относительно класса Φ .

5. Тесты для классов Поста

Тестирование неисправностей для булевских функций особо важно в связи с их прикладной значимостью.

Здесь мы остановимся на тестировании одиночных константных неисправностей для этих функций, но не только во всем классе P_2 , как мы это делали в § 4, а для подклассов функций из P_2 . В качестве классификации булевских функций берется классификация Э. Поста.

В 1921 г. Э. Поста решил проблему описания всех итеративно замкнутых классов булевских функций с явным указанием этих классов.

Здесь будет описано решение задачи указания значений функции Шеннона $L(n, F_{st}^2)$ для каждого из указанных классов Поста K в случае одиночных константных неисправностей. Эти функции будем обозначать для краткости $L_k(n)$.

Описываемые результаты принадлежат О. А. Долотовой [16].

Опишем классы Поста. Следуя обозначениям Поста, обозначим класс всех булевских функций через C_1 .

Функция $f^*(x_1, \dots, x_n) = f(\bar{x}_1, \dots, \bar{x}_n)$ называется двойственной к $f(x_1, \dots, x_n)$ из C_1 и называется самодвойственной, если $f^*(x_1, \dots, x_n) = f(x_1, \dots, x_n)$.

Функция $f(x_1, \dots, x_n)$ называется монотонной, если для любых двух наборов $\alpha = (a_1, \dots, a_n)$ и $\beta = (b_1, \dots, b_n)$ из того, что для всех $i = 1, 2, \dots, n$ выполнено $a_i \leq b_i$, то есть $\alpha \leq \beta$, следует, что $f(\alpha) \leq f(\beta)$.

Функция $f(x_1, \dots, x_n)$ называется линейной, если $f(x_1, \dots, x_n) = \sum_{i=1}^n c_i x_i + d \pmod{2}$.

Говорят, что функция из C_1 удовлетворяет условию $\langle a^\mu \rangle$, $\mu \geq 2$, если любые μ наборов, на которых она равна нулю, имеют общую нулевую компоненту; функция удовлетворяет условию $\langle a^\infty \rangle$, если все наборы, на которых она равна нулю, имеют общую нулевую компоненту.

Выражения $\bigvee_{j=1}^n x_j$ и $\bigwedge_{j=1}^m x_j$ называются, соответственно, логическими суммами и произведениями. Функция $f(x_1, \dots, x_n)$ называется α -функцией, если $f(x, \dots, x) = x$ и β -функцией, если $f(x, \dots, x) = 1$.

Опишем классы Поста.

Класс A_1 состоит из всех монотонных функций; D_3 — всех самодвойственных функций; F_4^μ — всех функций со свойством $\langle a^\mu \rangle$; F_4^∞ — всех функций со свойством $\langle a^\infty \rangle$; L_1 — всех линейных функций; C_2, A_2, L_2 — из всех монотонных функций, соответственно, из классов C_1, A_1, L_1 ; D_2, F_3^∞, F_3^μ — из всех монотонных функций, соответственно, из классов D_3, F_4^∞, F_4^μ ; $C_4, A_4, D_1, L_4, F_1^\mu, F_2^\mu, F_1^\infty, F_2^\infty$ —

из всех α -функций, соответственно из классов $C_1, A_1, D_3, L_3, F_1^\mu, F_4^\mu, F_3^\infty, F_4^\infty; L_4$ — из всех линейных самодвойственных функций; $C_3, A_3, L_3, F_1^\infty, F_1^\mu$ — из всех функций, двойственных ко всем функциям, соответственно, из классов $C_2, A_2, L_2, F_1^\infty, F_1^\mu, \mu = 2, 3, \dots$. Класс S_1 состоит из всех логических сумм: $S_3 = S_1 \cup \{1\}, S_5 = S_1 \cup \{0\}, S_6 = S_1 \cup \{0, 1\}$.

Класс P_1 состоит из всех логических произведений, $P_3 = P_1 \cup \{1\}, P_5 = P_1 \cup \{0\}, P_6 = P_1 \cup \{0, 1\}$.

Классы S_i, P_i при $i = 1, 3, 5, 6, L_j$ при $j = 1, 2, 3, 4, 5, D_l$ при $l = 1, 2, 3, A_m, C_m$ при $m = 1, 2, 3, 4, F_s^\infty, F_s^\mu, s = 1, 2, \dots, 8, \mu = 2, 3, \dots$, называем, соответственно, классами типов S, P, L, D, A, C и F .

Теорема 19. *Имеют место следующие соотношения:*

- 1) $L_k(n) = 2$ для каждого класса K типа L ;
- 2) $L_k(n+1) = 2$ для каждого класса K типов S или P ;
- 3) $L_k(n) \sim 2n$ при $n \rightarrow \infty$ для каждого класса K типов D, M, C или F .

Обозначим через $\bar{L}_k(n)$ функцию Шеннона для почти всех функций из K .

Теорема 20. *Имеют место следующие соотношения:*

- 1) $\bar{L}_k(n) = 2$ для каждого класса K типа L ;
- 2) $\bar{L}_k(n) = n+1$ для каждого класса K типов S или P ;
- 3) $\bar{L}_{D_2}(n) = 4$ при n четном; $\bar{L}_{D_2}(n) \in \{4, 6\}$ при n нечетном;
- 4) $\bar{L}_k(n) \in \{2, 3, 4\}$ для классов D_1, D_2 ;
- 5) $\bar{L}_k(n) = 4$ для каждого класса K типа A ;
- 6) $\bar{L}_k(n) = 3$ для каждого класса K типа C ;
- 7) $\bar{L}_k(n) \in \{4, 5\}$ для каждого класса $K \in \{F_1^\infty, F_4^\infty, F_5^\infty, F_8^\infty, F_1^\mu, F_4^\mu, F_5^\mu, F_8^\mu\}$ при $\mu = 3, 4, \dots$;
- 8) $3 \leq \bar{L}_k(n) \leq 8$ для каждого класса $K \in \{F_1^2, F_4^2, F_5^2, F_8^2\}$;
- 9) $\bar{L}_k(n) = 5$ для каждого класса $K \in \{F_2^\infty, F_3^\infty, F_6^\infty, F_7^\infty, F_2^\mu, F_3^\mu, F_6^\mu, F_7^\mu\}$ при $\mu = 3, 4, \dots$

- 10) $\bar{L}_k(n) = 4$ при n нечетном, $\bar{L}_k(n) \in \{4, 5\}$ при n четном для каждого класса $K \in \{F_2^2, F_3^2, F_6^2, F_7^2\}$.

Особый интерес вызывает поведение длины минимального теста для конкретных видов функций из классов Поста.

Используя для краткости и уточнения записи $L(f, F_{st}^2)$ запись $L_k(f)$, указывая тем самым, что мы имеем дело с длиной минимального теста для f из K по отношению к одиночной константной неистинности.

Теорема 21. *Справедливы следующие положения:*

- 1) $L_k(f) = 2$ для любой функции f из классов K типа L , существенно зависящей от $n \geq 1$ переменных;
- 2) $L_k(f) = n + 1$ для любой функции f из классов типов S или P , существенно зависящей от n переменных;
- 3) если
 - а) $t(n) = 2n - 4(p + 1)$ при $2^p + 2p < n \leq 2^{p+1} + 2(p + 1)$;
 - б) $C_k = 2$ для каждого класса K типа C и классов D_1 и D_3 ;
 - в) $C_k = 3$ для каждого класса $K \in \{F_1^\infty, F_4^\infty, F_5^\infty, F_8^\infty, F_1^\mu, F_4^\mu, F_5^\mu, F_8^\mu\}$, $\mu = 2, 3, \dots$;
 - г) $C_k = 4$ для каждого класса K типа A , D_2 и классов $K \in \{F_2^2, F_3^2, F_6^2, F_7^2, F_2^3, F_3^3, F_6^3, F_7^3\}$;
 - д) $C_k = 5$ для каждого класса $K \in \{F_2^\infty, F_3^\infty, F_6^\infty, F_7^\infty, F_2^\mu, F_3^\mu, F_6^\mu, F_7^\mu\}$, $\mu = 4, 5, \dots$,

то для любых задач n или r таких, что $C_k \leq k \leq t(n)$, существует функция $a \in K$, существенно зависящая от n переменных, для которой $L_k(f) = r$, причем значение r для класса D_2 может быть только четным числом.

Изложенные факты могут быть представлены таблицей 1 со специальным комментарием. В ней представлено распределение величины $L_k(f)$ для функций, существенно зависящих от n переменных, из классов Поста.

Использованы следующие обозначения: если в r -ом столбце напротив соответствующих классов стоит знак:

- + — то существуют функции f , у которых $L(f) = r$, но они составляют бесконечно малую долю в классах;

r	1	2	3	4	5	6	7	8	9	10	...	n+1	n+2	...	-2n
$C_i(n)$ $i=1,2,3,4$	-	+	+	+	+	+	+	+	+	+	...	+	+	...	+
$L_i(n)$ $i=1,2,3,4,5$	-	+	-	-	-	-	-	-	-	-	...	-	-	...	-
$S_i(n), P_i(n)$ $i=1,3,5,6$	-	-	-	-	-	-	-	-	-	-	...	+	-	...	-
$M_i(n)$ $i=1,2,3,4$	-	-	-	+	+	+	+	+	+	+	...	+	+	...	+
$D_2(n)$ $n=$	2p	-	-	+	-	+	-	+	-	+	...	+	-	...	+
	2p+1	-	-	-	+	-	+	-	+	-	...	+	-	...	+
$D_1(n), D_3(n)$	-	+	+	+	+	+	+	+	+	+	...	+	+	...	+
$F_i^\omega(n), F_i^\mu(n)$ $i=2,3,6,7$ $\mu=4,5,...$	-	-	-	-	+	+	+	+	+	+	...	+	+	...	+
$F_i^3(n)$ $i=2,3,6,7$	-	-	-	+	+	+	+	+	+	+	...	+	+	...	+
$F_i^2(n)$ $n=$	2p	-	-	+	+	+	+	+	+	+	...	+	+	...	+
	2p+1	-	-	-	+	+	+	+	+	+	...	+	+	...	+
$F_i^\omega(n), F_i^\mu(n)$ $i=1,4,5,8$ $\mu=3,4,...$	-	-	+	+	+	+	+	+	+	+	...	+	+	...	+
$F_i^2(n)$ $i=1,4,5,8$	-	-	+	+	+	+	+	+	+	+	...	+	+	...	+

Таблица 1. Пояснительная таблица.

- - - то в классах не существует функции f , у которой $L(f) = r$;
- если знаки + или - из столбцов $r_i, r_{i+1}, \dots, r_{i+k}$ обведены:
- сплошной линией, то все или почти все функции f из классов имеют $r_i \leq L(f) \leq r_{i+k}$;

- пунктиром, то существенные доли функций f из классов имеют $r_i \leq L(f) \leq r_{i+k}$.

Эти результаты позволяют осуществить сравнительный анализ сложности контроля управляющих систем, соответствующих различным классам Поста: проблема контроля управляющих систем в рассматриваемой постановке для различных классов Поста имеет количественно и качественно различные решения. В то же время просматриваются общие закономерности. Для счетного множества классов Поста почти все функции из этих классов имеют минимальные проверяющие тесты, состоящие из ограниченного числа наборов, но в каждом из них хотя и не существует функций, зависящих от n переменных, сложность минимальных тестов которых была бы равна в точности $2n$, имеются функции, сложность минимальных тестов которых равна любому из промежуточных значений, начиная от некоторой определенной для каждого класса константы до величины, асимптотически равной $2n$, то есть числу возможных неисправностей рассматриваемого типа. Таким образом, в каждом из этих классов имеются функции, у которых почти все из числа $2n$ неисправностей проверяются отдельным набором из минимальных тестов каждая. Только конечное число классов не имеет подобной структуры по сложностям минимальных тестов классов функций: в классах линейных функций типа L все функции имеют минимальные тесты сложности 2, в классах логических сумм и произведений типа S или P сложность минимальных тестов всех функций f равна $n + 1$, где n — число существенных переменных этих функций. И, наконец, для класса D_2 — монотонных самодвойственных функций, характеристика сложности минимальных тестов отличается от описанной выше закономерности только тем, что функции из этого класса не могут иметь минимальных тестов, состоящих из нечетного числа наборов.

Список литературы

- [1] Чегис И. А., Яблонский С. В. Логические способы контроля электрических схем // Труды математического института им. В. А. Стеклова АН СССР. Т. 51. 1958. С. 270–360.

- [2] Кренделев Ф. П., Дмитриев А. Н., Журавлев Ю. И. Сравнение геологического строения зарубежных месторождений докембрийских конгломератов с помощью дискретной математики // Доклады АН СССР. Т. 173, 5. 1967.
- [3] Переяславский В. В. Кандидатская диссертация по мат. кибернетике. — Механико-математический факультет МГУ им. Ломоносова М. В. 1958. С. 1–80.
- [4] Шаеб А. Исследование свойств линейных метрических алгоритмов распознавания. Кандидатская диссертация по мат. кибернетике. — Механико-математический факультет МГУ им. Ломоносова М. В. 1958. С. 1–126.
- [5] Константинов Р. М., Королева З. Е., Кудрявцев В. Б. Комбинаторно-логический подход к задачам прогноза рудоносности // Проблемы кибернетики. Т. 31. М.: Наука, 1976. С. 25–41.
- [6] Нефидов Ф. Н. Построение модели лечения детей с некоторыми острыми заболеваниями брюшной полости. Кандидатская диссертация по мат. кибернетике. — ВЦ АН СССР. 1977. С. 1–143.
- [7] Журавлев Ю. И. Об алгебраическом подходе к решению задач распознавания и классификации // Проблемы кибернетики. Т. 31. М.: Наука, 1978. С. 5–68.
- [8] Слепян В. А. Параметры распределения тупиковых тестов и информационные веса столбцов в бинарных таблицах // Дискретный анализ. Т. 14. Новосибирск, 1969. С. 28–43.
- [9] Носков В. Н. О сложности тестов, контролирующей работу входов логических схем // Математические заметки. Т. 18. № 1. 1975. С. 137–150.
- [10] Дюкова Е. В. Асимптотически оптимальные тестовые алгоритмы в задачах распознавания. Кандидатская диссертация по мат. кибернетике. — ВЦ АН СССР. 1979. С. 1–127.
- [11] Андреев А. Е. О качественных и метрических свойствах тестовых алгоритмов. Кандидатская диссертация по мат. кибернетике. — Механико-математический факультет МГУ им. Ломоносова М. В. 1981. С. 1–145.

- [12] Кибкало А. А. О T -алгоритмах распознавания, использующих короткие тесты. Кандидатская диссертация по мат. кибернетике. — Механико-математический факультет МГУ им. Ломоносова М. В. 1988. С. 1–153.
- [13] Кузнецов В. Е. Об одном стохастическом алгоритме вычисления информационных характеристик таблиц по методу тестов // Дискретный анализ. Т. 23. СОАН СССР. Новосибирск: Наука, 1973. С. 8–23.
- [14] Носов М. В. Функциональные характеристики тестовых алгоритмов распознавания образов. Кандидатская диссертация по мат. кибернетике. — Механико-математический факультет МГУ им. Ломоносова М. В. 1989. С. 1–80.
- [15] Погосян Г. Р. О сложности проверяющих тестов для логических устройств. Кандидатская диссертация по мат. кибернетике. — ВЦ АН СССР. 1982. С. 1–73.
- [16] Долотова О. А. О сложности контроля логических схем типа Поста. Кандидатская диссертация по мат. кибернетике. — ВЦ АН СССР. 1991. С. 1–101.
- [17] Соловьев Н. А. Тесты (теория, построение, применение). Монография. Новосибирск: Наука, 1978.
- [18] Яблонский С. В., Гаврилов Г. П., Кудрявцев В. Б. Функции алгебры логики и классы Поста. Монография. М.: Наука, 1966. С. 1–98.