Тестовое распознавание*

В. Б. КУДРЯВЦЕВ, А. Е. АНДРЕЕВ
Московский государственный университет
им. М. В. Ломоносова
e-mail: andreev@lsi.com

УДК 519.95

Ключевые слова: тест, распознавание, тупиковый тест, короткий тест, минимальный тест, оценка сложности.

Аннотация

В работе описывается логический подход к распознаванию образов. Ключевым понятием логического подхода является тест. Анализ тестов позволяет давать функциональную характеризацию образов, а также строить процедуры для вычисления. Рассматриваются качественные и количественные характеристики тестов, функционалов и процедур распознавания. Приводятся решения ряда известных проблем.

Abstract

V. B. Kudryavtsev, A. E. Andreev, Test recognition, Fundamentalnaya i prikladnaya matematika, vol. 15 (2009), no. 4, pp. 67—99.

We describe the logic approach to pattern recognition; its key notion is a test. Analyzing the tests allows us to construct functional characterizing the pattern, as well as procedures to compute them. We present qualitative and quantitative properties of tests, functionals, and recognition procedures. Solutions of a series of known problems are also given.

Введение

Многие задачи распознавания могут быть описаны с помощью следующей схемы. Имеется некоторый объект A, который может находиться в состояниях q_i, $i \in \{1, 2, \ldots, m\}$. Последние характеризуются параметрами-признаками x_j, $j \in \{1, 2, \ldots, n\}$. Известно, что некоторые наборы значений этих признаков $\alpha = (a_1, a_2, \ldots, a_n)$ описывают состояние q_i. Эти наборы как строки составляют матрицу T_i. Совокупность T (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 набор не содержится в них, то он не соответствует никакому состоянию. Если он входит в подматрицу T_i, то в предположении, что подматрицы не пересекаются, этот набор описывает состояние q_i.

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

Фундаментальная и прикладная математика, 2009, том 15, № 4, с. 67—99.

В реальной ситуации и признаки, и сами подматрицы T_i, как правило, описываются не полностью, а потому решение вопроса о том, какое состояние описывает данный набор, становится нетривиальным. Этот вопрос и называют задачей распознавания.

Задачу распознавания можно сформулировать так: даны матрица T' и её 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, породившие соответствующие направления в теории распознавания. В них накоплен большой опыт, включающий как перечень решённых и потенциально решаемых задач, так и методов их решения.

Особый класс составляют задачи, в которых характеризация 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 $\alpha = (a_1, a_2, \ldots, a_n)$ можно, как отмечалось, решить задачу распознавания состояния q объекта A.

С. В. Яблонский и И. А. Чегис [16] заметили, что, вообще говоря, такое распознавание можно осуществлять, используя не всю матрицу T, а только её часть. Ими было введено понятие теста для T.

Пусть σ — набор признаков, $T_{i,\sigma}$ — часть таблицы T_i, образованная столбцами, соответствующими σ, 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 $\Im(T)$ всех тестов для T. Этот подход нашёл применение в технической диагностике. Позже он был распространён 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' «мало» отличаются друг от друга.

Ю. И. Журавлёв, Ф. П. Кренделев и А. Н. Дмитриев [7] предложили оценивать роли признаков в решении задачи распознавания как долю их вхождения в $\Im(T)$. Величина p_j для признака x_j, называемая информационным весом, позволила им рассмотреть при специальной кодировке значений признаков линейный функционал $\sum\limits_{j=1}^{n} p_j x_j$, ввести некоторые пороги d_i для T_i и по значению $\sum\limits_{j=1}^{n} p_j x_j$ и d_i определять близость набора значений признаков $\alpha = (a_1, a_2, \ldots, a_n)$

к некоторой из матриц T_i. На этом пути ими было предложено решение ряда задач по оценке месторождений полезных ископаемых.

Позже В. И. Переяславским [12] был исследован вопрос, когда линейный функционал $\sum\limits_{j=1}^{n} p_j x_j$ доставляет верное решение задачи распознавания. К этим исследованиям примыкают рассмотрения А. Шайеба [17], посвящённые выяснению возможностей линейных функционалов в решении задачи распознавания в общем случае.

Другое развитие идеи использования тестов было осуществлено авторами. Как отмечалось, тест σ может выступать в качестве «эксперта» для определения по набору α состояния q объекта A. В случае когда $\sigma \in \Im(T')$, он по α из T уже, вообще говоря, не решает задачи принадлежности α некоторому T_i, поскольку $\Im(T')$ может не совпадать с $\Im(T)$. 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 $\Im(T')$. Используя каждый тест σ из $\Im(T')$ для α, получаем вектор «голосов» $\chi(\alpha) = (k_1, \ldots, k_m, k_{m+1})$, где k_i — число тестов, высказавшихся за T_i при $i = 1, 2, \ldots, m$, а k_{m+1} — число «воздержавшихся».

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

Функционал χ был успешно опробован в решении ряда прикладных задач, но вместе с тем обнаружились и определённые трудности его использования. 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 никакое собственное подмножество признаков не образует теста. 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, обозначим через $\chi_{\text{ТТ}}$, $\chi_{\text{КТ}}$ и $\chi_{\text{МТ}}$, а сами семейства назовём основными.

Выяснились огромные преимущества функционала $\chi_{\text{КТ}}$ в решении задач распознавания. С его помощью было решено большое число задач геологии, экономики, военного дела, а также задач из других областей. Были, например, установлены новые месторождения нефти, газа и олова в Сибири [6], оценены перспективы развития конкретных экономических районов, проводился анализ текущей военно-политической ситуации, устанавливался диагноз заболевания и оптимальный режим лечения для него [9].

Идея голосования была использована Ю. И. Журавлёвым и его учениками при голосовании по подмножествам признаков, состоящих из заданного числа элементов r. В частности, была предложена оптимизация выбора конкретного значения r, при котором распознавание оказывается наилучшим [4].

Важность функционалов χ, $\chi_{\text{ТТ}}$, $\chi_{\text{КТ}}$ и $\chi_{\text{МТ}}$ в решении задач распознавания, подтверждённая практикой, создала предпосылки для разработки тестовой теории распознавания. Её создание предполагало решение следующих задач:

1) получение оценок для числа основных видов тестов матрицы T, имеющей заданные параметры подматриц T_i;
2) нахождение точных и приближённых 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 задачи 3);
5) выделение из основных семейства, которое лучше остальных решает задачу распознавания для T';
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' состоит из одной строки, первые результаты были получены В. А. Слепян [14] и В. Н. Носковым [10]. Они нашли оценки сверху и снизу для числа тестов и тупиковых тестов в таких матрицах.

Затем в предположении, что в T признаков «много больше», чем строк, Е. В. Дюкова [3] разработала специальную комбинаторную технику, с помощью которой она нашла асимптотики тестов и тупиковых тестов для почти всех таблиц T' с заданными соотношениями параметров её размерностей; эта техника позволила ей синтезировать оптимальные детерминированные и стохастические алгоритмы построения тестов и тупиковых тестов, а также изучить поведение весов признаков и их корреляции. Таким образом, Е. В. Дюковой удалось частично решить задачи 1), 2), 6), 7), 9). Работа Е. В. Дюковой была прорывом в теории тестового распознавания.

Достижения Е. В. Дюковой затем были развиты и существенно усилены А. Е. Андреевым [1]. Им была разработана новая специальная техника, которая позволила

- 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) путём указания возможных доопределений матриц T'_i таким образом, чтобы множества $\Im(T')$ и $\Im(T)$ 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 определённой 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 идёт распознавание. Этому вопросу были посвящены исследования А. А. Кибкало [5]. 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 тест с числом признаков, «близким» к $\log m - \log\log m$.

А. А. Кибкало для случая, когда T' состоит из двух подматриц, установил следующее:

- 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 растёт 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 (т. е. близких к минимальным тестам) тестов перед остальными в выявлении признаков, лучше других отличающих подматрицы T'_1 и T'_2. А. А. Кибкало существенно продвинулся в решении задачи 7).

А. А. Кибкало также рассматривал ситуацию, близкую к задачам 3) и 5), когда в матрице T'

- 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, т. е. 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.

Исторически первый приближённый алгоритм решения задач 6) и 7) был предложен В. Е. Кузнецовым [8]. Этот алгоритм типа Монте-Карло использовался в процедурах решения практических задач распознавания. Затем приближённые алгоритмы для этих задач построили Е. В. Дюкова, А. Е. Андреев и А. А. Кибкало. Позже М. В. Носовым [11] был предложен алгоритм вычисления χ, основанный на экспоненциально сложном анализе множества T' с последующим полиномиальным по сложности вычислением χ по набору α.

Восстанавливая хронологию, необходимо подчеркнуть, что первые задачи, которые удалось решить Ю. И. Журавлёву и его коллегам, имели матрицу T' с числом строк и столбцов до 20. Работа В. Е. Кузнецова позволила решать подобные задачи для T' до 100 строк и столбцов. Е. В. Дюкова расширила этот класс до матриц, имеющих 200 столбцов и 50 строк, а А. Е. Андреев — до матриц размером в 200 строк и 200 столбцов. А. А. Кибкало смог использовать уже матрицы размером в 500 строк и 500 столбцов. Современные компьютеры позволяют увеличить размеры реально обрабатываемых матриц на порядок и более.

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 следующей модели.

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

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

Отметим здесь две группы основных результатов.

Первую составляют результаты Г. Р. Погосяна [13], который описал поведение $L_K(r, \gamma)$ для γ, соответствующих коротким замыканиям, слипаниям, инверсиям и др. Ранее в этом направлении интересные продвижения были получены В. Н. Носковым [10].

Вторую группу составляют результаты О. А. Долотовой [2], которая для тех же типов ошибок нашла поведение функции для $L_2^C(r, \gamma)$, аналогичной функции $L_C(r, \gamma)$, но соотнесённой с заданным классом Поста C функций алгебры логики. К этим группам результатов примыкают подобные рассмотрения 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 [15].

Статья, помимо введения, содержит пять разделов. В первом разделе приводятся тестовые результаты для матриц T' с малым числом строк. Во втором разделе излагаются результаты по тестам в общем случае при произвольном графе сравнения подматрицы T'. В третьем разделе рассматриваются короткие тесты. В четвёртом разделе содержатся основные факты по тестированию схем, реализующих функции k-значной логики. В пятом разделе приводятся результаты по тестированию схем, реализующих функции из классов Поста.

1. Тесты для матриц с малым числом строк

В этом разделе мы излагаем основные результаты Е. В. Дюковой по задачам 6) и 7).

Пусть матрица T имеет m строк и n столбцов, а её элементы a_{ij} входят в $E_k = \{0, 1, \ldots, k-1\}$, $k \geqslant 2$. Пусть $m = \sum\limits_{i=1}^{s} m_i$, где m_i — натуральные числа, $m_i \leqslant 1$, $\tilde{m}$ — последовательность $m_1, m_2, \ldots, m_s$. Пусть $M^{(k)}_{\tilde{m},n,s}$ — 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 $l \in \{1, 2, \ldots, s\}$ полагаем $m''(l) = \sum\limits_{i=1}^{l} m_i$, $m'(l) = m''(l) - m_l + 1$. Считаем, что в матрице из $M^{(k)}_{\tilde{m},n,s}$ подматрица T'_l содержит строки с номерами $m'(l), m'(l+1), \ldots, m''(l)$. Пусть $\Im_{\text{т}}(T')$ — множество всех тупиковых тестов для T'. Для строк $\alpha = (a_{j_1}, a_{j_2}, \ldots, a_{j_n})$ и $\alpha' = (a_{j'_1}, a_{j'_2}, \ldots, a_{j'_n})$ из T' введём операцию $\oplus$ сложения по модулю два, полагая

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

Пусть $h = \sum\limits_{l=1}^{s-1} \sum\limits_{t=1}^{s-l} m_l m_{l+t}$. Составим матрицу $L_{T'}$ с элементами b_{uv}, $u \in \{1, 2, \ldots, h\}$, $v \in \{1, 2, \ldots, n\}$, из всех таких строк $\alpha \oplus \alpha'$, что строки α и α' входят в разные подматрицы T'_i и $T'_{i'}$, причём $i < i'$. Полагаем строки в $L_{T'}$ упорядоченными следующим образом: если α, α', α'', α''' имеют в T номера j_1, j_2, l_1, l_2, 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 $\Re = \{R_1, R_2, \ldots, R_g\}$, где R_W — некоторый набор элементов из $L_{T'}$, $W = \{1, 2, \ldots, g\}$. Требуется построить все подмножества R, для которых выполнен ряд свойств $B_1, B_2, \ldots, B_p$, $p \geqslant 2$.

Искомую совокупность обозначим через $Q_{\Re}$. Предполагаем, что между $Q_{\Re}$ и $\Im_{\text{т}}(T')$ существует взаимно-однозначное соответствие, так что построение $Q_{\Re}$ приводит к построению $\Im_{\text{т}}(T')$.

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

Пример 1. Пусть $g = u$ и элемент $R_W \in \Re$ является столбцом матрицы $L_{T'}$ с номером W. Для наборов столбцов матрицы L_T рассмотрим 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 найдётся столбец, пересечение которого с этой строкой даёт 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 $\Im_{\text{т}}(T')$. В самом деле, набор σ номеров столбцов $j_1, j_2, \ldots, j_r$ матрицы T' является тупиковым тестом в точности 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 $L_{T'}$ назовём совместимым, если выполнено хотя бы одно из следующих 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 $\Re$ является элементом матрицы $L_{T'}$ с номером $W \in \{1, 2, \ldots, g\}$ и $\Omega(D)$ — набор тех столбцов из $L_{T'}$, в которых расположены элементы множества $D \subseteq R$. Будем говорить, что D обладает свойством B_1', если $D \in S(L_{T'})$. Если $D = \{b_{u_1v_1}, b_{u_2v_2}, \ldots, b_{u_rv_r}\}$, будем говорить, что D обладает свойством B_2', если выполнены следующие условия:

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

Множество $D \subseteq \Re$ 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' $\sigma = (j_1, j_2, \ldots, j_r)$ и множество $D \in \mathcal{B}(L_{T'})$ считаем эквивалентными, если $\Omega(D)$ состоит из столбцов матрицы $L_{T'}$ с номерами $j_1, j_2, \ldots, j_r$. Можно показать, что между $\mathcal{B}(L_{T'})$ и $\Im_{\text{т}}(T')$ имеется взаимно-однозначное соответствие, при котором соответствующие элементы эквиваленты. Следовательно, построение $\mathcal{B}(L_{T'})$ приводит к построению $\Im_{\text{т}}(T')$.

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

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

Алгоритм $\mathcal{A}$ назовём асимптотически оптимальным в $N_{m,n}$, если для почти всех подматриц T' из $N_{m,n}$ выполнено $\mu(\mathcal{A}, T') \sim |\Im_{\text{т}}(T')|$ при $m, n \to \infty$. Будем предполагать, что число столбцов в матрице T' «много» больше числа строк.

Нетрудно убедиться, что примером алгоритма для построения $\Im_{\text{т}}(T')$, не являющимся асимптотически оптимальным, является процедура из [3], которую мы обозначим через $\mathcal{A}_1$. Алгоритм $\mathcal{A}_1$ сводит задачу построения $\Im_{\text{т}}(T')$ к задаче, описанной в примере 1, и для построения погружения использует свойство B_1. Величина $\mu(\mathcal{A}, T')$, равная числу тестов в T', почти всегда по порядку больше величины $|\Im_{\text{т}}(T')|$.

Сведение задачи построения $\Im_{\text{т}}(T')$ к задаче из примера 2 позволяет построить асимптотически оптимальный по сложности алгоритм. Этот алгоритм $\mathcal{A}_2$ для построения погружения использует свойство B_1', и для него $\mu(\mathcal{A}, T') = |S(T')|$. Следовательно, вопрос об асимптотической оптимальности $\mathcal{A}_2$ сводится к оценке величин $|S(T')|$ и $|\Im_{\text{т}}(T')|$.

Справедливо следующее утверждение.

Теорема 1. *Если*

$$\alpha > 1, \quad \beta < \frac{1}{2}, \quad (m_1, m_2)^\alpha \leqslant n \leqslant k^{(m_1,m_2)^\beta}, \quad \varphi_k = (a, b),$$

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 $m, n \to \infty$ для почти всех T из M_m выполнено

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

Из теоремы 1 следует, что при

$$(m_1 m_2)^\alpha \leqslant n \leqslant k^{(m_1 m_2)^\beta}, \quad \alpha > 1, \quad \beta < \frac{1}{2},$$

алгоритм $\mathcal{A}_1$ является асимптотически оптимальным в $F^{(k)}_{m_1,m_2,n}$.

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

Следует отметить, что асимптотика величины $|\Im_{\text{т}}(T')|$ для почти всех матриц T' из $M^{(2)}_{\tilde{m},n,s}$ в случае, когда $s = m$, найдена в [10, 14], где при получении оценки для $|\Im_{\text{т}}(T')|$ так же, как и в доказательстве теоремы 1, используется то, что почти всегда величина $|\Im_{\text{т}}(T')|$ асимптотически совпадает со своим средним значением.

Наряду с детерминированной процедурой построения множества $\Im_{\text{т}}(T')$ Е. В. Дюковой были получены стохастические алгоритмы, основанные на алгоритме $\mathcal{A}_2$.

При стохастическом подходе используется не всё множество тупиковых тестов матрицы T', а лишь случайная выборка из него. Анализ этой выборки даёт возможность приближённо решать задачи быстрого распознавания, выяснения роли отдельных признаков в задаче распознавания и корреляции между ними, т. е. задачи 6) и 7), а также оценивать при этом возможную ошибку.

Так как объём выборки существенно меньше множества всех тупиковых тестов, то при этом удаётся значительно увеличить размеры обрабатываемых таблиц.

В упомянутом ранее алгоритме В. Е. Кузнецова [8] построение тупиковых тестов для T' из $M^{(k)}_{\tilde{m},n,s}$ осуществляется по следующей схеме.

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

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 $\Im_{\text{т}}(T')$.

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

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

Алгоритм $\mathcal{A}_2$ быстро работает на таблицах, в 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 $u = \{l_1, l_2, \ldots, l_i\}$ определяет подматрицу $L_{T'}(u)$ матрицы $L_{T'}$, образованную строками с номерами $l_1, l_2, \ldots, l_i$. Матрица $L_{T'}(u)$ обрабатывается при помощи алгоритма $\mathcal{A}_1$, который строит $\Im(T', u)$. Обработка некоторой совокупности случайных наборов из U_i приводит к построению случайной выборки из $\Im_{\text{т}}(T')$.

Эта схема в деталях отличается от схемы В. Е. Кузнецова, но при нахождении меры важности признака оказывается, что при определённых условиях оба подхода почти всегда эквивалентны детерминированному подходу. Пусть $\Im_{t,j}(T')$ — множество всех тупиковых тестов, содержащих признак x_j, а $\Im_j(T', u)$ — множество тупиковых тестов в $\Im(T', u)$, содержащих признак x_j. В дополнение к информационному весу введём следующие величины:

$$p(x_j) = \frac{|\Im_{t,j}(T')|}{|\Im(T')|}, \qquad p^i(x_j) = \frac{1}{C_n^i} \sum_{u \in U_i} \frac{|\Im_i(T', u)|}{|\Im(T', u)|},$$

$$\bar{p}^i(x_j) = \frac{\sum\limits_{u \in U_i} |\Im_i(T', u)|}{\sum\limits_{u \in U_i} |\Im(T', u)|}, \qquad \bar{p}^i(x_j, u) = \frac{|\Im_i(T', u)|}{|\Im(T', u)|},$$

где $u \in U_i$.

Теорема 2. *Если $u \in U_i$, $n \leqslant k^{(m_1 m_2)^\beta}$, $\beta < \frac{1}{3}$ и при $m, n \to \infty$ справедливо*

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

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

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

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

В этом разделе мы излагаем основные результаты А. Е. Андреева по задачам 1), 2), 6), 8), 9), которые существенно усиливают результаты Е. В. Дюковой.

Введённые ранее понятия теста, тупикового теста и др. в более общей ситуации, которая рассматривается здесь, нуждаются в уточнениях. С этой целью введём необходимый формализм.

Для множеств A и B положим

$$A^B = \{f \mid f \colon B \to A\}.$$

Если $f \in A^B$ и $C \subseteq B$, то

$$f(C) = \{b \mid b = f(a),\ a \in C\},$$

а если $a \in A$ и $C \subseteq A$, то

$$f^{-1}(a) = \{b \mid f(b) = a\}, \quad f^{-1}(C) = \bigcup_{a \in C} f^{-1}(a).$$

Пусть

$$\langle A^B \rangle = \{f \mid f \in A^B,\ f(b) = a\}.$$

Если $f \in A^B$, $g \in B^C$, то $f \circ g \in A^C$ и $(f \circ g)(x) = f\big(g(x)\big)$. Пусть $A \otimes B$ — множество упорядоченных пар (a, b), где $a \in A$, $b \in B$, $A \times B$ — множество неупорядоченных пар $\{(a, b)\}$, причём $a \neq b$. 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 положим $1_x^A(y) = 1$, 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 $\tilde{\varepsilon}^A = \{1_x^A \mid x \in A\}$, $\varepsilon^A = \tilde{\varepsilon}^A \cup \{1^h\}$. Знаки $\leqslant$ и $<$ для функций из E^A используем в обычном смысле для указания поэлементной упорядоченности на этом множества. Если $a, b \in E^A$, положим

$$a * b = a \oplus b \oplus 1^A, \quad |a| = |a^{-1}(1)|, \quad \rho(a, b) = |a \oplus b| = |A| - |a * b|.$$

При $m \geqslant 1$ положим $L_m = \{1, 2, \ldots, m\}$. 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 $a(i)$. Теперь мы можем рассматривать упорядоченные наборы длины n элементов из A как отображения из A^{L_m}. Пусть $E^* = \bigcup\limits_{n=1}^{\infty} E^n$, $E_r^n = \{x \mid x \in E^n,\ |x| = r\}$. Положим

$$1^n = 1^{L_n}, \quad 0^n = 0^{L_n}, 1_i^n = 1_i^{L_n}, \quad 0_i^n = 0_i^{L_n}, \quad i \in L_n, \tilde{\varepsilon}^n = \tilde{\varepsilon}^{L_n}, \quad \varepsilon^n = \varepsilon^{L_n}.$$

Пусть $a \in E^n$. Определим отображение η_a из $L_n^{L_{|a|}}$, положив $\eta_a(i) = k$, если $\sum\limits_{j=1}^{k-1} a(j) < i$, где $i = \sum\limits_{j=1}^{k} a(j)$. Эта функция интерпретируется следующим образом. 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 определим отображение P_r^a из $(E^{|a|})^{E_n}$, полагая, что $P_r^a(b) = b \circ \eta_a$.

Введём ряд обозначений, связанных с графами. Пусть $V(G)$ и $X(G)$ — множества вершин и рёбер графа G соответственно. Предполагаем, что $X(G) \subseteq V(G)^{(2)}$. Пусть $\gamma_G(x)$ — степень вершины x, $\Delta(G)$ — максимальная степень вершины, $\chi(G)$ — хроматическое число, $d(r, G)$ — число остовных r-рёберных лесов. Через $k(G)$ обозначим число компонент связности, а через $k^*(G)$ — число нетривиальных компонент, т. е. компонент, содержащих не менее двух вершин.

Пусть

$$p(G) = |V(G)|, \quad q(G) = |X(G)|, \quad p^*(G) = p(G) - k(G) + 1.$$

Если U — конечное множество и $M = \{V_1, V_2, \ldots, V_k\}$ — система его подмножеств, то M называется покрытием для U, если $\bigcup\limits_{i=1}^{k} V_i = U$.

Через $G(M)$ обозначим граф с множеством вершин U и множеством рёбер $U^{(2)} \setminus \Big(\bigcup\limits_{i=1}^{k} V_i^{(2)} \Big)$.

Назовём 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 таких пар (T, G), что $T \in (E^n)^U$. Обобщим ранее введённое понятие теста до понятия G-теста, но для краткости сохраним за ним название «тест».

Пусть заданы конечное множество U, его покрытие $M = \{V_1, V_2, \ldots, V_k\}$ и граф G с множеством вершин U. Набор a из E_2^n назовём тестом пары (T, M) из $\Psi_{M,n}$, если для любых a_1, a_2 из U, таких что $T(a_1) * T(a_2) \geqslant a$, найдётся i из L_k, для которого $a_1, a_2 \in V_i$. Набор a из E_2^n назовём тестом пары (T, G) из $\Psi_{G,n}$, если ни для каких смежных вершин a_1, a_2 графа G не выполнено $T(a_1) * T(a_2) \geqslant a$. Набор a из E_2^n назовём тестом матрицы T из $(E^n)^U$, если ни для какого b из U не выполнено $T(b) \geqslant a$.

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

Для T из $(E^n)^U$ определим $U^{(2)}$-матрицу $T^{(2)}$, положив $T^{(2)}(\{a_1, a_2\}) = T(a_1) * T(a_2)$. Для (T, G) из $\Psi_{G,n}$ положим $T_G = T^{(2)}\big|_{X(G)}$ и назовём эту матрицу матрицей сравнения пары (T, G). Для пары (T, M) из $\Psi_{M,n}$, где T — U-матрица, M — покрытие множества U, положим $T_M = T_{G(M)}$ и назовём T_M 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 — U-матрица, M — покрытие U*;
2) *набор a является тестом* (*тупиковым тестом*) *пары* $\big(T, G(M)\big)$;
3) *набор a является тестом* (*тупиковым тестом*) *таблицы T_M*.

Пусть $\mathbb{N}$, $\mathbb{R}$ и $\mathbb{R}_+$ — множества натуральных, действительных и положительных действительных чисел соответственно, $\mathcal{M}$ — множество таких непрерывных отображений α из $\mathbb{R}_+^{\mathbb{R}_+}$, что $\lim\limits_{x \to \infty} \alpha(x) = 0$,

$$\mathcal{B} = \left\{ D \mid D = \frac{1}{\alpha},\ \alpha \in \mathcal{M} \right\}.$$

Знаки $\leqslant_n^{\circ}$, $\geqslant_n^{\circ}$, $=_n^{\circ}$ используются для обозначения неравенств и равенств, которые выполняются начиная с некоторого n_0. Точнее, если существует α из M, такое что $B \leqslant_n^{\circ} \big(1 + \alpha(n)\big) A$, то пишем $B \lesssim_n A$ (при $n \to \infty$ асимптотически не превосходит); если $B \lesssim_n A$ и $A \lesssim_n B$, то пишем $A \sim_n B$ (при $n \to \infty$ асимптотически равны). Если выполнено $c_1 A \leqslant B \leqslant c_2 A$ (либо $c_1 A \leqslant_n^\circ B \leqslant_n^\circ c_2 A$), где $c_1, c_2 \in \mathbb{R}_+$, то пишем $B \asymp A$ (или $B \asymp_n A$). Если для $\alpha \in M$ верно, что $A \leqslant_n^\circ \alpha(n) B$, то пишем $A \ll_n B$.

Будем говорить, что при некоторых ограничениях на пары (G, n) некоторое свойство выполнено для почти всех пар (T, G) из $\Psi_{(G,n)}$ при $n \to \infty$, если существует α из M, такое 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 $B_{G,n}$, определённых на $\Psi_{(G,n)}$, то пишем $A_{G.n} \sim_n^{\text{п. в.}} B_{G.n}$ (это означает, что существуют α и β из M, такие что доля пар (T, G) из $\Psi_{(G,n)}$, для которых не выполнено $A_{G.n}(T,G)\bigl(1-\beta(n)\bigr) \leqslant B_{G.n}(T,G) \leqslant \bigl(1+\beta_n(n)\bigr)A_{G.n}(T,G)$ не превосходит $\alpha(n)$). Аналогично определяются обозначения $\leqslant_n^{\text{п. в.}}$, $=_n^{\text{п. в.}}$, $\geqslant_n^{\text{п. в.}}$ и т. д.

Значения функций $\varphi_{G,n}^{\text{тт}}$ и $\varphi_{G,n,r}^{\text{тт}}$ на матрице T равны соответственно числу тупиковых тестов и числу тупиковых тестов длины r пары (T, G). Далее через $\log x$ будем обозначать $\log_2 x$. Пусть $\chi(x) = \frac{\log x - \log\log x}{2}$, $H(m, r) = \mathrm{e}^{-m \cdot 2^{-r}}\bigl(1 - \mathrm{e}^{-mr \cdot 2^{-r}}\bigr)^r$. Для действительного x и натуральных m и n, таких что $0 \leqslant x \leqslant n$, положим $\hat{H}(n,m,r) = \binom{n}{x} H(m,r)$. Для $0 \leqslant x \leqslant n$ положим $\binom{n}{x} = \frac{\Gamma(n+1)}{\Gamma(x+1)\Gamma(n-x+1)}$, где Γ — гамма-функция, в противном случае считаем, что $\binom{n}{x} = 0$. Если существует, и притом единственное, решение уравнения $\hat{H}(n,m,x+1) = \hat{H}(n,m,x-1)$ относительно x из интервала $(1, n-1)$, положим $r(n,m)$ равным этому решению, иначе положим $r(n,m) = 1$.

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

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

Положим

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

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

Пусть

$$\Phi^2_{D,\varepsilon}(n,m) = \sum_{r=r^1_{D,\varepsilon}(n,m)}^{r^2_{D,\varepsilon}(n,m)} \binom{n}{r} H(m,r).$$

Положим

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

если $q(G) \leqslant n^{1-\varepsilon}$;

$$\tilde r^1_{D,\varepsilon}(G,n) = r^1_{D,\varepsilon}\big(n,q(G)\big), \quad \tilde r^2_{D,\varepsilon}(G,n) = r^2_{D,\varepsilon}\big(n,q(G)\big),$$

если $q(G) > n^{1-\varepsilon}$. Пусть

$$\Phi_{D,\varepsilon}(G,n) = \begin{cases} \Phi^1_D(G,n), & \text{если} \;\; q(n) \leqslant n^{1-\varepsilon}, \\ \Phi^1_{D,\varepsilon}\big(n,q(G)\big), & \text{если} \;\; q(n) > n^{1-\varepsilon}. \end{cases}$$

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

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

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

$$\varphi^{\text{тт}}_{G,n} \sim^{\text{п. в.}}_n \sum_{r=\tilde r^1_{D,\varepsilon}(G,n)}^{\tilde r^2_{D,\varepsilon}(G,n)} \varphi^{\text{тт}}_{G,n,r} \sim^{\text{п. в.}}_n \Phi_{D,\varepsilon}(G,n);$$

2) *если $q(G) \geqslant 2^n D(n)$, то*

$$\varphi^{\text{тт}}_{G,n} =^{\text{п. в.}}_n 0;$$

3) *не существует единственной функции f, определённой на множестве пар (G,n), где G — граф и $n \in \mathbb{N}$, для которой выполнено, что если $c_1 2^n \leqslant q(G) \leqslant c_2 2^n$, то $\varphi^{\text{тт}}_{G,n} \sim^{\text{п. в.}}_n f(G,n)$ (или $\varphi^{\text{тт}}_{G,n} =^{\text{п. в.}}_n f(G,n)$).*

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

1) $L^{\min}_{G,n}(T) = n+1$, если (T,G) не имеет тестов;
2) $L^{\min}_{G,n}(T) = \min\{r \mid \varphi^{\text{тт}}_{B,n,r}(T) \neq 0\}$ 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

$$(1-\varepsilon)\lambda + \lambda\log l + (1-\lambda)\log(1-\lambda) = 0, \quad D \in B, \;\; D(n) \leqslant^{\circ}_n \log\log n.$$

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

1) $]\log\chi(G)[$, если $p^*(G) \leqslant (\ln n)\log\ln n$ или если $(\ln n)\log\ln n < p^*(G) \leqslant (\log n)^{D(n)}$ и $\chi(G) < 2\big((\ln n)\log\ln n\big)^{3/4}$;
2) $\big[S_1\big(\binom{\chi(G)}{2}, n\big) + \varepsilon\big]$, если $(\ln n)\log\ln n < p^*(G) \leqslant (\log n)^{D(n)}$ и $\chi(G) \geqslant 2\big((\ln n)\log\ln n\big)^{3/4}$;

3) $[S(\Delta(G), n) + \varepsilon]$, если $(\log n)^{D(n)} < p^*(G)$, $q(G) < 2^{n-\lambda_\varepsilon}$ и $\Delta(G) > \frac{(1-\varepsilon)q(G)(\ln q(G))}{\ln\binom{n}{\log q(G)}}$;
4) $[S(q(G), n) + \varepsilon]$, если $(\log n)^{D(n)} < p^*(G)$, $q(G) < 2^{n\cdot\lambda_\varepsilon}$ и $\Delta(G) \leqslant \frac{(1-\varepsilon)q(G)(\ln q(G))}{\ln\binom{n}{\log q(G)}}$ или если $(\log n)^{D(n)} < p^*(G)$ и $2^{n-\lambda_\varepsilon} \leqslant q(G) < \frac{2^n}{(\log n)^{D(n)}}$.

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

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

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

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

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

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

Следующее утверждение описывает поведение длин минимальных тестов.

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

1) *если $1 \leqslant q(G) \leqslant \frac{2^n}{(\log n)^{D(n)}}$, то*
$$L^1_{D,\varepsilon}(G, n) \leqslant^{\text{п. в.}}_n L^{\min}_{D,n} \leqslant^{\text{п. в.}}_n L^2_{D,\varepsilon}(G, n);$$
2) *если $(\log n)^{D(n)} \leqslant q(G) \leqslant \frac{2^n}{(\log n)^{D(n)}}$ и $q(G) \geqslant p^*(G)^{1+\varepsilon}$, то*
$$[S(q(G), n) + \varepsilon] \leqslant^{\text{п. в.}}_n L^{\min}_{b,n} \leqslant^{\text{п. в.}}_n [S(q(G), n) + \varepsilon] + 1;$$
3) *если $q(G) \geqslant \frac{2^n}{(\log n)^{D(n)}}$, то*
$$\hat{L}^1_{D,\varepsilon}(G, n) \leqslant^{\text{п. в.}}_n \hat{L}^{\min}_{D,n} \leqslant^{\text{п. в.}}_n \hat{L}^2_{D,\varepsilon}(G, n).$$

Опишем процедуру построения тупиковых тестов. Положим $L_{m,n} = L_m \otimes L_n$. Матрицей T с m строками и n столбцами называем матрицу, введённую ранее, а также отображение $T\colon L_{m,n} \to E$. Положим $\Psi_{m,n} = E^{L_{m,n}}$, $\Psi^* = \bigcup\limits_{(m,n)\in\mathbb{N}^2} \Psi_{m,n}$. Если $T \in \Psi_{m,n}$, то через $\hat{T}$ обозначим матрицу из $\Psi_{L_m,n}$, такую что $\big(\hat{T}(k)\big)(l) = T(k,l)$ для любой пары (k,l) из $L_{m,n}$. Набор x назовём тестом или тупиковым тестом матрицы T из $\Psi_{m,n}$, если он является тестом или соответственно тупиковым тестом матрицы $\hat{T}$. Положим $\Lambda_{m,n} = (E^m \setminus \{0^m\}) \otimes (E^n \setminus \{0^n\})$.

Каждой паре (a,b) из $\Lambda_{m,n}$ соответствует подрешётка $a^{-1}(1) \otimes b^{-1}(1)$ решётки $L_{m,n}$. 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 $\big(T \circ (\eta_a, \eta_b)\big)(i,j) = T\big(\eta_a(i), \eta_b(j)\big)$. Считаем, что $T^{(a,b)} = T|_{a^{-1}(1) \otimes b^{-1}(1)}$. Набор x из E^n, $x \leqslant b$, 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 её фрагментов, просматриваемых алгоритмом в процессе работы.

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

1) $\{1^m\} \otimes (E_2^n \setminus \{0^n\}) \subseteq \Sigma(m,n)$, $(m,n) \in \mathbb{N}^2$;
2) если $F(T) = 1$, то T — тупиковая тестовая матрица, т. е. набор из всех единиц является её единственным тупиковым тестом;
3) если $(a,b) \in \Sigma(m,n)$ и x — тупиковый тест подматрицы $T(a,b)$, то существует c, $c \leqslant a$, такое что
$$(c,x) \in \Sigma(m,n), \quad F\big(T^{(c,x)}\big) = 1.$$

Схема $A = (\Sigma, F)$ называется локальной, если для любой пары (a,b) из $\Sigma(m,n)$ верно, что если $c \leqslant a$, $x \leqslant b$, то $(c,x) \in \Sigma(m,n)$ в точности тогда, когда $(c \circ \eta_a, x \circ \eta_b) \in \Sigma(|a|,|b|)$.

Введём граф $G_A(T)$ с множеством вершин
$$V_A(T) = \big\{(a,b) \mid (a,b) \in \Sigma(m,n),\ F\big(T^{(a,b)}\big) = 1\big\} \cup \{(0^m, 0^n)\}$$
и множеством рёбер $X_A(T)$, 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 $a_1 < a < a_2$, $(a,b_1) \in \Sigma(m,n)$, 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.

Введём 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 будем писать $A \leqslant B$, когда для любой матрицы T из Ψ^* верно $\mu(A,T) \leqslant \mu(B,T)$. Аналогично определяется и неравенство $A \leqslant^* B$.

Этим требованиям удовлетворяет, например, схема алгоритма Е. В. Дюковой, о которой говорилось в разделе 1. Там $A^* = (\Sigma^*, F^*)$, $\Sigma^*(m,n) = \Lambda_{m,n}$, $F^*(T) = 1$ в точности тогда, когда T — квадратная тупиковая тестовая матрица. Схема A^* связная и локальная.

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

$$1^k = C_f^{k,k} > C_f^{k,k-1} > \ldots > C_f^{k,1} > C_f^{k,0} = 0^k,$$
$$(C_f^{k,p} - C_f^{k,p-1})\Big(\eta_{C_f^{k,p}}\big(f(p)\big)\Big) = 1.$$

Определим схему алгоритма $D(f)$ следующим образом:

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

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

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

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 $\hat{S}^{-1}(1) = S(L_l)$. Текущие параметры работы D-алгоритма с матрицей T из $\Psi_{m,n}$ «в момент t», $t \in \mathbb{N}$, — пятёрка (t, K_t, S_t, l_t, A_t), где $l_t \in \mathbb{N}$, $K_t \in L_m^{L_{l_t}}$, $S_t \in L_n^{L_{l_t}}$, $A_t \in \Psi_{l_t,n}$.

1. Если $\hat{T}(1) \in 1^n$, то T не имеет тестов. Конец.
2. Положим

$$t = 1, \quad l_t = 1, \quad K_t(1) = 1,$$
$$\hat{A}_t(1) = \hat{T}(1) \oplus 1^n, \quad S_t(1) = \min\big(\hat{T}(1)\big)^{-1}(0).$$

   Перейти к п. 3.
3. Положим $\varepsilon = \big(\hat{T}^{(1^m, \tilde{S}_t)}\big)^{-1}(1^{l_t})$. Если $\varepsilon = \varnothing$, $\tilde{S}_t$ — тупиковый тест, перейти к п. 5; если $\varepsilon \neq \varnothing$, перейти к п. 4.
4. Положим $k = \min \varepsilon$, $\varepsilon_i = \big(\hat{T}^{(C_{f_0}^{m,k}, \tilde{S}_t)}\big)^{-1}(1_i^{l_t})$,

$$y = \bigg(\mathop{\&}_{i=1}^{l_t}\Big(\bigvee_{r \in \varepsilon_i} \hat{T}(r)\Big)\bigg) \& (\hat{T}(k) \oplus 1^n).$$

   Если $y = 0^n$, перейти к п. 5. В противном случае увеличим t на 1 и получим

$$l_t = l_{t-1} + 1, \quad \hat{A}_t(l_t) = y, \quad k_t(l_t) = k,$$
$$(K_t, S_t, \hat{A}_t)|_{L_{l_{t-1}}} = (K_{t-1}, S_{t-1}, \hat{A}_{t-1}), \quad S_t(l_t) = \min y^{-1}(1).$$

Перейти к п. 3.

5. Положим $C_r = \big(\hat{A}_t(r)\big)^{-1} \cap \big(L_n \setminus L_{S_t(r)}\big)$. Если для всех r из L_{l_t} верно $C_r = \varnothing$, то конец. В противном случае положим $j_0 = \min C_{r_0}$, где $r_0 = \min\{r \mid C_r \neq \varnothing\}$, увеличим t на 1 и положим
$$S_t(l_t) = \min y^{-1}(1), \quad S_t|_{L_{r_0-1}} = S_{t-1}|_{L_{r_0-1}},$$
$$S_t(r_0) = j_0, \quad l_t = r_0, \quad (K_t, \hat{A}_t) = (K_{t-1}, \hat{A}_{t-1})|_{L_{r_0}}.$$
Перейти к п. 3. Если π — нумерация рёбер 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 $T_G \circ \pi^{-1}$. Положим $\mu\big((A, \pi), (T, G)\big) = \mu(A, T_G \circ \pi^{-1})$ и аналогично для μ^*. Здесь матрицы T и $\hat{T}$ отождествляются.

Теорема 7. *Если каждому графу G соответствует нумерация его рёбер π_G и $\log q(G) \ll_n \log^2 n$, то*
$$\mu\big((D, \pi_G), (T, G)\big) = \mu^*\big((D, \pi_G), (T, G)\big) \lesssim_n^{\text{п. в.}} \varphi^{\text{тт}}_{G,n}(T, G).$$

3. Короткие тесты

В этом разделе 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, мы рассматриваем 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 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.

Эти множества мы будем называть опорными.

Для 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 весов по всем тупиковым тестам, растут с ростом доли пар строк из разных подматриц, различаемых признаком. Такие тесты считаются короткими.

Пусть Ω — оператор, ставящий в соответствие таблице T' опорное множество тестов $\Omega(T')$. Оператор Ω, а также алгоритм распознавания, основанный на $\Omega(T')$, считается устойчивым, если отношение $\frac{|\Omega(T')\cap\Omega(T^*)|}{|\Omega(T')\cup\Omega(T^*)|}$ близко к единице, если таблицы T и T^* различаются мало. Для оператора Ω типа а) почти все наборы признаков являются тестами, поэтому он, как отмечалось выше, не представляет интереса.

Для типов б), в) и г) 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' несёт 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.

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, $i=1,2,\ldots$, и число столбцов у T_1 и T_2 равно n. Обозначим через $\hat{T}$ таблицу сравнения для T.

Пусть $\varphi^{\text{т}}_{v_1,v_2,n,\tilde{x}}$, где $\tilde{x}\in E_2^n\setminus\{\tilde{o}^n\}$ — функция, определённая на множестве $\Psi_{v_1,v_2,n}$ и принимающая значение 1, если $\tilde{x}$ является тестом для T, и значение 0 иначе.

Для r через $\varphi^{\text{т}}_{v_1,v_2,n,\tilde{x}}$ обозначим функцию, равную числу тестов длины r у T из $\Psi_{v_1,v_2,n}$, а через $\varphi^{\text{т}}_{v_1,v_2,n}$ 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

$$\varphi^{\text{т}}_{v_1,v_2,n,r}=\sum_{\tilde{x}\in E_r^n}\varphi^{\text{т}}_{v_1,v_2,n,\tilde{x}},\varphi^{\text{т}}_{v_1,v_2,n}=\sum_{i=1}^{n}\varphi^{\text{т}}_{v_1,v_2,n,r}.$$

Аналогично определяется функция $\varphi^{\text{тт}}_{v_1,v_2,n,\tilde{x}}$, которая равна 1, если $\tilde{x}$ — тупиковый тест для T, и 0 в остальных случаях. Определим функцию $\varphi^{\text{тт}}_{v_1,v_2,n,r}$ равной числу тупиковых тестов длины r, $\varphi^{\text{тт}}_{v_1,v_2,n}$ равной числу всех тупиковых тестов, $\varphi^{\text{кт}}_{v_1,v_2,n,r}$ равной числу тестов длины не более r, $\varphi^{\text{ктт}}_{v_1,v_2}$ равной числу всех тупиковых тестов длины не более r.

Для этих функций выполнено

$$\varphi^{\text{тт}}_{v_1,v_2,r} = \sum_{\tilde{x}\in E_2^{n,r}} \varphi^{\text{тт}}_{v_1,v_2,\tilde{x}}, \quad \varphi^{\text{тт}}_{v_1,v_2,r} = \sum_{r=1}^{n} \varphi^{\text{тт}}_{v_1,v_2,n,r},$$

$$\varphi^{\text{кт}}_{v_1,v_2,n,r} = \sum_{j=1}^{r} \varphi^{\text{т}}_{v_1,v_2,n}, \quad \varphi^{\text{ктт}}_{v_1,v_2,n,r} = \sum_{j=1}^{r} \varphi^{\text{тт}}_{v_1,v_2,n,j}.$$

Если $\nabla \in \{\text{т}, \text{тт}\}$, $\Delta \in \{\text{т}, \text{тт}, \text{кт}, \text{ктт}\}$, $i \in L_n$, $r \in L_n$, положим

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

$$\varphi^{\nabla,i}_{v_1,v_2,n} = \sum_{r=1}^{n} \varphi^{\nabla,i}_{v_1,v_2,r}, \quad \varphi^{\text{к},\nabla,i}_{v_1,v_2,n,r} = \sum_{j=1}^{r} \varphi^{\nabla,i}_{v_1,v_2,n,j}.$$

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

$$\psi^{\Delta,i}_{v_1,v_2,r} = \frac{\varphi^{\Delta,i}_{v_1,v_2,n,r}}{\varphi^{\Delta}_{v_1,v_2,n,r}}, \quad \psi^{\nabla,i}_{v_1,v_2,n} = \frac{\varphi^{\nabla,i}_{v_1,v_2,n}}{\varphi^{\nabla}_{v_1,v_2,n}}.$$

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

- $\psi^{\text{т},i}_{v_1,v_2,n}$ — вес по всем признакам;
- $\psi^{\text{тт}}_{v_1,v_2,n}$ — вес по всем тупиковым тестам;
- $\psi^{\text{т},i}_{v_1,v_2,n,r}$ — вес по всем тестам длины r;
- $\psi^{\text{кт},i}_{v_1,v_2,n,r}$ — вес по всем тестам длины не более r;
- $\psi^{\text{ктт},i}_{v_1,v_2,n,r}$ — 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 $\eta_{\tilde{x}}\colon L_{|\tilde{x}|} \to L_n$, $\eta_{\tilde{x}}(i) = k_i$, $i \in L_{|\tilde{x}|}$, где k_i — номер i-й единицы набора $\tilde{x}$.

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

Для $\tilde{x}$ из E_2^n определим отображение $\pi_{\tilde{x}}\colon E_2^n \to E_2^{|\tilde{x}|}$ 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 $\pi_{\tilde{x}}$ обозначим $\pi_{n,l}$.

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

$$\varepsilon m_i \leqslant |(\pi_{n,1} \circ T_i)^{-1}(1)| \leqslant (1-\varepsilon)m_i, \quad i \in L_2, \quad |(\pi_{n,1} \circ T_i)^{-1}(1)| = \alpha m_1 m_2.$$

Рассматриваются такие α, что множество $\Psi^{\alpha,\varepsilon}_{v_1,v_2,n}$ 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 мы обозначим I^{ε}.

Через $\psi^{\alpha,i}_{v_1,v_2,n}$, $\psi^{\alpha,\text{тт}}_{v_1,v_2,n}$, $\psi^{\alpha,\text{т}}_{v_1,v_2,n,r}$, $\psi^{\alpha,\text{тт}}_{v_1,v_2,n,r}$, $\psi^{\alpha,\text{кт}}_{v_1,v_2,n,r}$, $\psi^{\alpha,\text{ктт}}_{v_1,v_2,n,r}$ обозначим сужение соответствующих весовых функций на множество $\Psi^{\alpha,\varepsilon}_{v_1,v_2,n}$.

Пусть

$$r^{\varepsilon}_{1,k}(m) =]\log m - (1-\varepsilon)\log\log m[,\quad r^{\varepsilon}_{2,k}(m) = [\log m - (1+\varepsilon)\log\log\log m],$$
$$\tilde r^{\varepsilon}_{1,k}(m) = [\log m - (1+\varepsilon)\log\log m].$$

Теорема 8. *Пусть c_1, c_2, ε — константы, $0<c_1<1<c_2$, $\varepsilon \in \big(0, \min(\frac{1}{32}, \frac{c_1}{20c_2})\big)$. Тогда если $n^{c_1} \leqslant m_1m_2 \leqslant n^{c_2}$, $m_1 \asymp_n m_2$, $\alpha_1, \alpha_2 \in I^{\varepsilon}$, то*

1) *существует такая функция $\psi^{\alpha}(n, m_1m_2)$, что*
$$\psi^{\alpha,\text{тт}}_{v_1,v_2,n} \sim \psi^{\alpha}(n, m_1m_2),$$
а при $\alpha_1 < \alpha_2$ выполнено
$$\psi^{\alpha_1}(n, m_1m_2) \ll_n \psi^{\alpha_2}(n, m_1m_2);$$
2) $\psi^{\alpha,\text{т}}_{v_1,v_2,n} \sim^{\text{п. в.}}_n \frac{1}{2}$;
3) *существует функция $\psi^{\alpha}_k(n, m_1m_2, r)$, для которой при $r \in [\tilde r^{\varepsilon}_{1,k}(m_1m_2), r^{\varepsilon}_{2,k}(m_1m_2)]$ выполнено*
$$\psi^{\alpha,\text{кт}}_{v_1,v_2,n,r} \sim^{\text{п. в.}}_n \psi^{\alpha,\text{ктт}}_{v_1,v_2,n,r} \sim^{\text{п. в.}}_n \psi^{\alpha,\text{т}}_{v_1,v_2,n,r} \sim^{\text{п. в.}}_n$$
$$\sim^{\text{п. в.}}_n \psi^{\alpha,\text{тт}}_{v_1,v_2,n,r} \sim^{\text{п. в.}}_n \psi^{\alpha}_k(n, m_1m_2, r);$$
если $\alpha_1 < \alpha_2$ и $r \in [\tilde r^{\varepsilon}_{1,k}(m_1m_2), r^{\varepsilon}_{2,k}(m_1m_2)]$, то
$$\psi^{\alpha_1}_k(n, m_1m_2, r) \gg_n \psi^{\alpha_2}_k(n, m_1m_2, r);$$
если $\alpha_1 < \frac{1}{2}$ и $r \in [\tilde r^{\varepsilon}_{1,k}(m_1m_2), \tilde r^{\varepsilon}_{2,k}(m_1m_2)]$, то
$$\psi^{\alpha}_k(n, m_1m_2, r) \sim_n 1.$$

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 неё вытекают следующие утверждения.

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) растёт 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, т. е. 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.

Определим множество

$$\Psi^2_{v_1,v_2,n} = \Psi_{v_1,v_2,n} \otimes \Psi_{v_1,v_2,n}$$

и рассмотрим конечное вероятностное пространство с распределением

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

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

$$\tilde\rho(T,T^*) = \sum_{i=1}^{r}\sum_{a\in V_i}\rho\bigl(T_i(a),T^*_j(a)\bigr).$$

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

Определим на $\Psi^2_{v_1,v_2,n}$ функции $\psi^{\text{тт},\&}_{v_1,v_2,n}$, $\psi^{\text{тт},\vee}_{v_1,v_2,n}$, $\psi^{\text{тт},\&}_{v_1,v_2,n,r}$, $\psi^{\text{тт},\vee}_{v_1,v_2,n,r}$, $\psi^{\text{ктт},\&}_{v_1,v_2,n,r}$, $\psi^{\text{ктт},\vee}_{v_1,v_2,n,r}$, $\psi^{\text{т},\&}_{v_1,v_2,n,r}$, $\psi^{\text{т},\vee}_{v_1,v_2,n,r}$, $\psi^{\text{кт},\&}_{v_1,v_2,n,r}$, $\psi^{\text{кт},\vee}_{v_1,v_2,n,r}$ так, что функции со значком & задают число наборов из E_2^n, являющихся тестами соответствующего вида как для правильно заданной, так и для искажённой матрицы, а функции со значком $\vee$ задают число наборов, являющихся тестами хотя бы одной из этих таблиц.

Множество тестов одного из указанных видов называется устойчивым к искажениям, если

$$\varphi^{*,\&} \sim^{\text{п. в.}}_n \varphi^{*,\vee},$$

где φ^* — функция, задающая число тестов соответствующего вида.

Теорема 9. *Если* $\varepsilon \in \bigl(0, \min(\frac{1}{32}, \frac{c_1}{20c_2})\bigr)$, $0 < c_1 < 1 < c_2$, $n^{c_1} \leqslant m_1m_2 \leqslant n^{c_2}$, $(\log m_1m_2)^{-2} \ll p \ll (\log m_1m_2)^{-(1+\varepsilon)}$ *при* $n\to\infty$, $\nabla \in \{\text{т}, \text{тт}, \text{кт}, \text{ктт}\}$, *то*

1) *при* $r =]\log m_1m_2 - \varepsilon\log\log m_1m_2[$ *или* $r =]\log m_1m_2 - b\log\log m_1m_2[$, $b > 1+\varepsilon$ *выполнено почти всегда при* $n\to\infty$

$$\psi^{\nabla,\&}_{v_1,v_2,n,r} \sim^{\text{п. в.}}_n \psi^{\nabla,\vee}_{v_1,v_2,n,r} \sim^{\text{п. в.}}_n (C^r_n)e^{(-\frac{m_1m_2}{2^r})};$$

2) $\psi^{\text{тт},\&}_{v_1,v_2,n} \ll^{\text{п. в.}}_n \psi^{\text{тт},\vee}_{v_1,v_2,n}$;

3) *при* $r =]\log m_1m_2 - a\log\log m_1m_2[$, $a \in (1+\varepsilon, 2-\varepsilon)$ *почти всегда при* $n\to\infty$ *выполнено*

$$\psi^{\nabla,\&}_{v_1,v_2,n,r} \ll^{\text{п. в.}}_n \psi^{\nabla,\vee}_{v_1,v_2,n,r}.$$

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 искажённой таблицы, в то время как множество коротких тестов практически не изменяется. Кроме того, множество тестов (тупиковых тестов), длины которых близки к минимальной, 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

$$\pi\colon x(G_{v_1,v_2}) \to L_{m_1 m_2}.$$

Пусть $\hat{T}_\pi = \hat{T} \circ \pi^{-1}$ и $\Psi^*_{m,n}$ — множество матриц S, $S\colon L_m \to E^n$. Положим $\Psi^* = \bigcup\limits_{(m,n)\in\mathbb{N}^2} \Psi^*_{m,n}$. Пусть $S \in \Psi^*$ и A — алгоритм построения тестов (тупиковых тестов). Под сложностью работы алгоритма $\mu^{\text{т}}(A,S)$ ($\mu^{\text{тт}}(A,S)$) понимается, как и ранее при описании результатов А. Е. Андрееева, 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 $S\colon L_l \to L_n$. Пусть $\hat{s} \in E^n$, $\hat{s}^{-1}(1) = S(L_l)$ и $\hat{T}^{\hat{s}}_\pi = \hat{T}_\pi \circ \pi_{\hat{s}}$. Доопределим матрицу $\hat{T}_\pi$ на L_{m+1} так, что $\hat{T}_\pi(m+1) = \tilde{o}^n$. 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 $S_t\colon L_{l_t} \to L_n$, тестом матрицы $\hat{T}_\pi$. Набор K_t, $K_t\colon L_{l_t} \to L_{m+1}$, и матрицы A_t, M_t из $\Psi^*_{l_t,n}$ служат для организации работы алгоритма.

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

1. Если $\hat{T}_\pi(1) = \tilde{1}^n$, то $\hat{T}_\pi$ не имеет тестов, перейти к п. 7. В противном случае перейти к п. 2.
2. Положим $t = 1$, $l_t = 1$, $K_t(1) = 1$, $A_t(1) = \hat{T}_\pi(1) \oplus \tilde{1}^n$, $M_t(1) = \tilde{o}^n$, $S_t(1) = \min A_t(1)^{-1}(0)$. Перейти к п. 3.
3. Пусть ε равно $(\hat{T}^{\hat{s}_t}_\pi)^{-1}(\tilde{1}^{l_t})$, если $K_t(l_t) < m+1$, и $m+1$ в противном случае. Если $\varepsilon \cap L_m = \varnothing$, то $\hat{s}_t$ — тест. Если $l_t = r-1$, перейти к п. 5. В ином случае перейти к п. 4.
4. Пусть $k = \min\varepsilon$ и $\tilde{y} = \big(\hat{T}_\pi(k) \vee M_t(l_t)\big) \oplus \tilde{1}^n$. Если $\tilde{y} = \tilde{o}^n$, перейти к п. 6. В противном случае положим $t = t+1$, $l_t = l_{t-1}+1$, $A_t(l_t) = \tilde{y}$, $S_t(l_t) = \min \tilde{y}^{-1}(1)$, $K_t(l_t) = K$, $M_t(l_t) = M_{t-1}(l_t - 1)$, $(K_t, S_t, A_t, M_t)|_{L_{l_{t-1}}} = (K_{t-1}, S_{t-1}, A_{t-1}, M_{t-1})$, перейти к п. 3.

5. Положим $\tilde{y} = \Big(\big(\bigvee\limits_{i\in\varepsilon} \hat{T}(i)\big) \vee M_t(l_t)\Big) \oplus \tilde{1}^n$. Для любого $i \in \tilde{y}^{-1}(1)$ объект $\hat{s}_t(+)\tilde{o}_j^n$ — тест. Положим $t = t + |\tilde{y}|$,
$$(l_t, S_t, A_t, M_t) = (l_{t-|\tilde{y}|}, S_{t-|\tilde{y}|}, K_{t-|\tilde{y}|}, M_{t-|\tilde{y}|}).$$
Перейти к п. 6.
6. Положим для $i \in L_{l_t}$
$$c_j = A_t(j)^{-1}(1) \cap \big(L_n \setminus L_{S_t(j)}\big).$$
Если $c_j = \varnothing$ для всех j на L_{l_t}, перейти к п. 7. В противном случае положим $t = t+1$, $l_t = \max\{j \mid c_j \neq \varnothing\}$, $M_t(l_t) = M_{t-1}(l_t) \vee \tilde{o}^n_{S_{t-1}(l_t)}$, $S_t(l_t) = \min c_{l_t}$, $(S_t, M_t)|_{L_{l_{t-1}}} = (S_{t-1}, M_{t-1})|_{L_{l_{t-1}}}$, $(K_t, A_t) = (K_{t-1}, A_{t-1})|_{L_{l_t}}$. Перейти к п. 3.
7. Закончить работу.

Теорема 10. *Если π из P_{v_1,v_2} — произвольная нумерация рёбер графа G_{v_1,v_2}, T — пара матриц из $\Psi_{v_1,v_2,n}$, то справедливы следующие утверждения*:

1) *при* $c > 1$, $\log m \leqslant (\log m)^c$, $a \in (0, \frac{1}{c})$ *и* $r =]\log m - a\log\log m[$ *выполнено*
$$\mu^{\text{тт}}(D_1(r), \hat{T}_\pi) \sim_n^{\text{п. в.}} \mu^{\text{т}}(D_2(r), \hat{T}_\pi) \sim_n^{\text{п. в.}} \psi^{\text{кт}}_{v_1,v_2,n,r};$$
2) *при* $\varepsilon \in (0,1)$, $\log m \leqslant 2^{(\log n)^{1-\varepsilon}}$, $b \in (1, \frac{1}{1-\varepsilon})$, $r =]\log m - b\log\log\log m[$ *выполнено*
$$\mu^{\text{тт}}(D_1(r), \hat{T}_\pi) \sim_n^{\text{п. в.}} \mu^{\text{т}}(D_2(r), \hat{T}_\pi) \sim_n^{\text{п. в.}} \psi^{\text{ктт}}_{v_1,v_2,n,r} \sim_n^{\text{п. в.}} \psi^{\text{кт}}_{v_1,v_2,n,r}.$$

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 её входным воздействиям, что, конечно, весьма громоздко и требует больших усилий.

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

Основным видом внешнего воздействия для проверки правильности работы системы является подача входных сигналов и проверка выходных реакций. Иногда удаётся за счёт 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. Мы изложим результаты Г. Р. Погосяна, полученные в этом направлении.

Формализуем сказанное. Отождествим элемент с функцией $f\colon E_k^n \to E_k$, которую будем называть функцией k-значной логики. Такие функции обычно записываются в виде $f(x_1, x_2, \ldots, x_n)$, где переменные x_i, как и сама функция, принимают значения из E_k. Пусть $x^n = \{x_1, x_2, \ldots, x_n\}$ и $\mathbb{N}_n = \{1, 2, \ldots, n\}$. Класс всех таких функций обозначим через P_k^n и положим $P_k = \bigcup\limits_{n=1}^{\infty} P_k^n$.

Пусть $\varphi\colon E_k^n \to E_k^n$. Это отображение будем называть φ-ошибкой. Пусть Φ_k^n — класс всех φ-ошибок от n переменных и $\Phi_k = \bigcup\limits_{n=1}^{\infty} \Phi_k^n$. С каждой φ-ошибкой и f свяжем функцию $f_\varphi(x_1, x_2, \ldots, x_n) = f\big(\varphi(x_1, x_2, \ldots, x_n)\big)$, которую будем называть φ-искажением функции f.

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

Для нас важно установить, когда $f \neq f_{\varphi_i}$ при любых $i = 1, 2, \ldots, s$. Проблема решается с помощью понятия теста, введённого 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 $\varphi \in \Phi$ либо для всех наборов α из E_k^n выполнено $f(\alpha) = f\big(\varphi(\alpha)\big)$, либо для некоторого набора β из T выполнено $f(\beta) = f\big(\varphi(\beta)\big)$.

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

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

Тест $T \in M(f, \Phi)$ называется минимальным, если $|T| = L(f, \Phi)$. Введём функцию Шеннона для длины проверяющего теста, полагая

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

Введём класс неисправностей F_{c}^k (будем называть его классом констант) следующим образом:

$$F_{\text{c}}^k = \big\{\varphi_\gamma\colon \gamma \in E_{k+1}^n \setminus (k, k, \ldots, k),\ \forall\, \alpha \in E_k^n\ \big(\varphi_\gamma(\alpha) = (c_1 * a_1, \ldots, c_n * a_n)\big)\big\},$$

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

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

$$L(n, F_{\text{c}}^k) = \begin{cases} 2n - 2t, & \text{\textit{если}}\ \ k^{t-1} + t < n \leqslant k^t + t, \\ (2n - 2t) + 1, & \text{\textit{если}}\ \ n = k^{t-1} + t. \end{cases}$$

Кратностью ошибки $\varphi_\gamma \in F_{\text{c}}^k$ будем называть число координат набора $\gamma = (c_1, \ldots, c_n)$, отличных от k. Множество всех ошибок из F_{c}^k, имеющих кратность не более чем p, обозначим через $F_{\text{c}}^k(p)$.

Пусть $M \subseteq E_k$. Положим

$$F_M^k = \{\varphi_\gamma \in F_{\mathrm{c}}^k \colon \forall\, i \in \mathbb{N}_n \ (c_i \in M \cup \{k\})\},$$

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

Теорема 12. *Для любых $n \geqslant 1$, $k \geqslant 2$, $p \in \mathbb{N}_n$ справедливо*

$$L\big(n, F_{\mathrm{c}}^k(p)\big) = L(n, F_{\mathrm{c}}^k).$$

Теорема 13. *Для любых $n \geqslant 1$, $k \geqslant 2$, $\mu \in \mathbb{N}_n$ справедливо*

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

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

$$L(n, F_\mu^k) = L(n, F_{\mathrm{c}}^k).$$

Введём класс ошибок F_{st}^k. С каждой ошибкой $\varphi \in F_{\mathrm{st}}^k$ однозначно свяжем разбиение $S(\varphi)$ множества переменных x_n на непересекающиеся непустые подмножества $Z_1(\varphi), \ldots, Z_{q_\varphi}(\varphi)$, где $q_\varphi \in \mathbb{N}_{n-1}$. При этом для любого набора α из E_k^n положим $\varphi(\alpha) = \beta = (b_1, \ldots, b_n)$, где для каждого i из $\mathbb{N}_n$ выполнено $b_i = \max\{a_j \colon x_j \in Z_l(\varphi)\}$, если $x_j \in Z_l(\varphi)$, $l \in \mathbb{N}_{q_\varphi}$.

Теорема 15. *Для любых $n \geqslant 2$, $k \geqslant 2$, $\mu \in \mathbb{N}_n$ справедливо*

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

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

Теорема 16. *Для любых $n \geqslant 2$, $k \geqslant 2$ и $p \in \mathbb{N}_{n-1}$ справедливо*

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

В случае $k = 2$ получаем

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

где $\alpha + \sigma$ обозначает набор, получающийся из α и σ их покоординатным сложением по модулю 2. Кроме этого сложения, введём покоординатное 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 определим следующим образом:

$$F_{\mathrm{in}}^2 = J_n \setminus \{\varphi_{\tilde{o}}\}.$$

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

$$2 \cdot \left[\frac{n-1}{2}\right] + 1 \leqslant L(n, F_{\mathrm{in}}^2) \leqslant n.$$

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

$$F_{\mathrm{in}}^2(1) = \{\varphi_{\delta^1}, \ldots, \varphi_{\delta^n}\}.$$

Теорема 18. *Для любого* $n \geqslant 1$ *справедливо*

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

где t *определяется из соотношения* $2^{t-1} + t \leqslant n \leqslant 2^t + t$.

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

Оказывается, что для любой функции $f \in P_2^n$ и любого теста $T \in M(f, F_{\mathrm{c}}^2)$ справедливо, что $T \in M\big(f, F_{\mathrm{in}}^2(1)\big)$, поэтому

$$L\big(n, F_{\mathrm{c}}^2 \cup F_{\mathrm{in}}^2(1)\big) = L(n, F_{\mathrm{c}}^2).$$

Функция $f \in P_k^n$ называется инвариантной относительно некоторой ошибки φ, если для любого $\alpha \in E_k^n$ справедливо равенство $f(\alpha) = f\big(\varphi(\alpha)\big)$. 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 $\{F_{\mathrm{c}}^k, F_{\mathrm{st}}^k, F_{\mathrm{in}}^2\}$ почти все функции из P_k^n являются чувствительными относительно класса Φ.

5. Тесты для классов Поста

В связи с большим прикладным значением булевских функций тестирование неисправностей для них особенно важно. Мы остановимся на тестировании одиночных константных неисправностей для этих функций, но не во всём классе P_2, как мы делали это в разделе 4, а в подклассах функций из P_2. В качестве классификации булевских функций берётся классификация Э. Поста.

В 1921 г. Э. Пост решил проблему описания всех итеративно замкнутых классов булевских функций, явно указав эти классы. Мы опишем решение задачи указания значений функции Шеннона $L(n, F_{\mathrm{st}}^2)$ для каждого из классов Поста K в случае одиночных константных неисправностей. Соответствующие функции будем обозначать для краткости $L_k(n)$. Излагаемые результаты принадлежат О. А. Долотовой [2].

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

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

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

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

Говорят, что функция из C_1 удовлетворяет условию $\langle a^\mu \rangle$, $\mu \geqslant 2$, 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 $\bigvee\limits_{j=1}^{n} x_j$ и $\bigwedge\limits_{j=1}^{m} x_j$ называются соответственно логическими суммами и произведениями. Функция $f(x_1, \ldots, x_n)$ называется α-функцией, если $f(x, \ldots, x) = x$, и β-функцией, если $f(x, \ldots, x) = 1$.

Опишем классы Поста.

Класс A_1 составляют все монотонные функции; класс D_3 — все самодвойственные функции; класс F_4^μ — все функции со свойством $\langle a^\mu \rangle$; класс F_4^∞ — все функции со свойством $\langle a^\infty \rangle$; класс L_1 — все линейные функции; классы C_2, A_2, L_2 — все монотонные функции соответственно из классов C_1, A_1, L_1; классы D_2, F_3^∞, F_3^μ — все монотонные функции соответственно из классов D_3, F_4^∞, F_4^μ; классы C_4, A_4, D_1, L_4, F_1^μ, F_2^μ, F_1^∞, F_2^∞ — все α-функции соответственно из классов C_1, A_1, D_3, L_3, F_1^μ, F_4^μ, F_3^∞, F_4^∞; класс L_4 — все линейные самодвойственные функции; классы C_3, A_3, L_3, F_1^∞, F_1^μ — все функции, двойственные ко всем функциям соответственно из классов C_2, A_2, L_2, F_1^∞, F_1^μ, $\mu = 2, 3, \ldots$. Класс S_1 составляют все логические суммы, $S_3 = S_1 \cup \{1\}$, $S_5 = S_1 \cup \{0\}$, $S_6 = S_1 \cup \{0, 1\}$. Класс P_1 составляют все логические произведения, $P_3 = P_1 \cup \{1\}$, $P_5 = P_1 \cup \{0\}$, $P_6 = P_1 \cup \{0, 1\}$.

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^∞, F_s^μ при $s = 1, 2, \ldots, 8$, $\mu = 2, 3, \ldots$, 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* $n \to \infty$ *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 нечётном* n; $\bar{L}_{D_2}(n) \in \{4, 6\}$ *при чётном* n;
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* $\mu = 3, 4, \ldots$;
8) $3 \leqslant L_k(n) \leqslant 8$ *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* $\mu = 3, 4, \ldots$;
10) $\bar{L}_k(n) = 4$ *при нечётном* n, $\bar{L}_k(n) \in \{4, 5\}$ *при чётном* n *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.

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

Теорема 21. *Справедливы следующие утверждения*:

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

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

   В таблице представлено распределение величины $L_k(f)$ *для существенно зависящих от* n *переменных функций из классов Поста. Использованы следующие обозначения*: *если в* r*-м столбце напротив соответствующих классов стоит знак* $+$, *то существуют функции* f, *для которых* $L(f) = r$, *но они составляют бесконечно малую долю в классах*; *если стоит знак* $-$, *то в классах не существует функции* f, *для которой* $L(f) = r$. *Если знаки* $+$ *или* $-$ *из столбцов* $r_i, r_{i+1}, \ldots, r_{i+k}$ *обведены сплошной линией, то для всех или почти всех функций* f *из классов справедливо* $r_i \leqslant L(f) \leqslant r_{i+k}$, *если знаки обведены пунктиром, то для существенных долей функций* f *из классов справедливо* $r_i \leqslant L(f) \leqslant r_{i+k}$.

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 счётного

r	1	2	3	4	5	6	7	8	9	10	...	n+1	n+2	...	~2n
$C_i(n)$ i=1,2,3,4	–	+	$\boxed{+}$	+	+	+	+	+	+	+	...	+	+	...	+
$L_i(n)$ i=1,2,3,4,5	–	$\boxed{+}$	–	–	–	–	–	–	–	–	...	–	–	...	–
$S_i(n), P_i(n)$ i=1,3,5,6	–	–	–	–	–	–	–	–	–	–	...	$\boxed{+}$	–	...	–
$M_i(n)$ i=1,2,3,4	–	–	–	$\boxed{+}$	+	+	+	+	+	+	...	+	+	...	+
$D_2(n)$ n= 2p	–	–	–	$\boxed{+}$	–	+	–	+	–	+	...	+	–	...	+
2p+1	–	–	–	+	–	+	–	+	–	+	...	+	–	...	+
$D_1(n), D_3(n)$	–	+	+	+	+	+	+	+	+	+	...	+	+	...	+
$F_i^\infty(n), F_i^\mu(n)$ i=2,3,6,7 μ=4,5,...	–	–	–	–	$\boxed{+}$	+	+	+	+	+	...	+	+	...	+
$F_i^3(n)$ i=2,3,6,7	–	–	–	+	$\boxed{+}$	+	+	+	+	+	...	+	+	...	+
$F_i^2(n)$ n= 2p	–	–	–	+	+	+	+	+	+	+	...	+	+	...	+
2p+1 i=2,3,6,7	–	–	–	$\boxed{+}$	+	+	+	+	+	+	...	+	+	...	+
$F_i^\infty(n), F_i^\mu(n)$ i=1,4,5,8 μ=3,4,...	–	–	+	+	+	+	+	+	+	+	...	+	+	...	+
$F_i^2(n)$ i=1,4,5,8	–	–	+	+	+	+	+	+	+	+	...	+	+	...	+

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 определённой для каждого класса константы до величины, асимптотически равной $2n$, т. е. числу возможных неисправностей рассматриваемого типа. Таким образом, в каждом из этих классов имеются функции, у которых почти все из $2n$ неисправностей проверяются отдельным набором из минимальных тестов. 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. 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 нечётного числа наборов.

Литература

[1] Андреев А. Е. О качественных и метрических свойствах тестовых алгоритмов. Дис.... канд. физ.-мат. наук. — М., 1981.

[2] Долотова О. А. О сложности контроля логических схем типа Поста. Дис.... канд. физ.-мат. наук. — М., 1991.

[3] Дюкова Е. В. Асимптотически оптимальные тестовые алгоритмы в задачах распознавания. Дис.... канд. физ.-мат. наук. — М., 1979.

[4] Журавлёв Ю. И. Об алгебраическом подходе к решению задач распознавания и классификации // Проблемы кибернетики. — 1978. — Т. 31. — С. 5—68.

[5] Кибкало А. А. О T-алгоритмах распознавания, использующих короткие тесты. Дис.... канд. физ.-мат. наук. — М., 1988.

[6] Константинов Р. М., Королёва З. Е., Кудрявцев В. Б. Комбинаторно-логический подход к задачам прогноза рудоносности // Проблемы кибернетики. — 1976. — Т. 31. — С. 25—41.

[7] Кренделев Ф. П., Дмитриев А. Н., Журавлёв Ю. И. Сравнение геологического строения зарубежных месторождений докембрийских конгломератов с помощью дискретной математики // ДАН СССР. — 1967. — Т. 173, № 5.

[8] Кузнецов В. Е. Об одном стохастическом алгоритме вычисления информационных характеристик таблиц по методу тестов // Дискрет. анализ. — 1973. — Т. 23. — С. 8—23.

[9] Нефидов Ф. Н. Построение модели лечения детей с некоторыми острыми заболеваниями брюшной полости. Дис.... канд. физ.-мат. наук. — М., 1977.

[10] Носков В. Н. О сложности тестов, контролирующих работу входов логических схем // Мат. заметки. — 1975. — Т. 18, № 1. — С. 137—150.

[11] Носов М. В. Функциональные характеристики тестовых алгоритмов распознавания образов. Дис.... канд. физ.-мат. наук. — М., 1989.

[12] Переяславский В. В. Дис.... канд. физ.-мат. наук. — М., 1958.

[13] Погосян Г. Р. О сложности проверяющих тестов для логических устройств. Дис.... канд. физ.-мат. наук. — М., 1982.

[14] Слепян В. А. Параметры распределения тупиковых тестов и информационные веса столбцов в бинарных таблицах // Дискрет. анализ. — 1969. — Т. 14. — С. 28—43.

[15] Соловьев Н. А. Тесты (теория, построение, применение). — Новосибирск: Наука, 1978.

[16] Чегис И. А., Яблонский С. В. Логические способы контроля электрических схем // Тр. Мат. ин-та им. В. А. Стеклова. — 1958. — Т. 51. — С. 270—360.

[17] Шайеб А. Исследование свойств линейных метрических алгоритмов распознавания. Дис.… канд. физ.-мат. наук. — М., 1958.

[18] Яблонский С. В., Гаврилов Г. П., Кудрявцев В. Б. Функции алгебры логики и классы Поста. — М.: Наука, 1966.