На правах рукописи

Кукарцев Анатолий Михайлович

ЭФФЕКТИВНЫЕ АЛГОРИТМЫ АНАЛИЗА ДЖЕВОНС-ЭКВИВАЛЕНТНОСТИ ДАННЫХ

Специальность 05.13.17 – Теоретические основы информатики

АВТОРЕФЕРАТ
диссертации на соискание учёной степени
кандидата физико-математических наук

Красноярск 2017

Работа выполнена в Федеральном государственном бюджетном образовательном учреждении высшего образования «Сибирский государственный аэрокосмический университет имени академика М.Ф. Решетнёва» (СибГАУ), г. Красноярск.

Научный руководитель: доктор физико-математических наук, доцент **Кузнецов Александр Алексеевич**.

Официальные оппоненты: **Винокуров Сергей Федорович**, доктор физико-математических наук, профессор, Федеральное государственное бюджетное образовательное учреждение высшего образования «Иркутский государственный университет», профессор кафедры Алгебраических и информационных систем;

Чехонадских Александр Васильевич, доктор технических наук, доцент, Федеральное государственное бюджетное образовательное учреждение высшего образования «Новосибирский государственный технический университет», профессор кафедры Алгебры и математической логики.

Ведущая организация: Федеральное государственное бюджетное образовательное учреждение высшего образования «Томский государственный университет систем управления и радиоэлектроники» (ТУСУР).

Защита состоится «07» апреля 2017 г. в 11.00 часов на заседании диссертационного совета Д 212.099.22 на базе Сибирского федерального университета по адресу: 660074, г. Красноярск, ул. ак. Киренского, 26, аудитория УЛК 112.

С диссертацией можно ознакомиться в библиотеке и на сайте Сибирского федерального университета по адресу http://www.sfu-kras.ru/.

Автореферат разослан «___» февраля 2017 г.

Учёный секретарь
диссертационного совета Покидышева Людмила Ивановна

Общая характеристика работы

Актуальность темы и степень её разработанности. Цифровая информация представляется преимущественно двумя способами: комбинационным и функциональным. Первый рассматривает её в виде комбинации символов некоторого алфавита, а второй – в виде множества значений некоторой дискретной функции (как в алгоритмах сжатия JPEG), в качестве которой может выступать булева функция (далее – БФ). Любому бинарному вектору (далее – БВ), состоящему из нулей и/или единиц, можно поставить в соответствие некоторую БФ. Для этого важно однозначно упорядочить область определения БФ. Сама БФ может быть описана реализующими её функциональными элементами или формулой. Множества отрицаний и перестановок аргументов БФ образуют группу Джевонса (или гипероктаэдральную группу) и определяют действие над БФ. Указанное действие замечательно тем, что оно нейтрально, т.к. не затрагивает связей между функциональными элементами и не меняет формулы. Оно задаёт естественную эквивалентность БФ и, как следствие, джевонс-эквивалентность соответствующих им данных.

Работы, связанные с группой Джевонса и её приложениями в теории информации, компьютерных науках, а также генетике, ведутся с середины XX века, – как за рубежом, (L. Geissinger и D. Kinch[1], W. H. Gates и C. H. Papadimitriou[2], S. Hannenhalli и P. A. Pevzner[3] и др.), так и в России (А. В. Тарасов[4], С. Ф. Винокуров и А. С. Казимиров[5], Б. А. Погорелов и М. А. Пудовкина[6], Е. К. Алексеев и Е. К. Карелина[7] и др.). Перечислим некоторые важные проблемы в указанной предметной области:

– анализ джевонс-эквивалентности данных;

– поиск элементов группы, связывающих джевонс-эквивалентные данные.

Поэтому необходимы эффективные алгоритмы, т.е. такие, которые могут находить решение указанных проблем за разумное время. Как известно, порядок группы Джевонса для БФ n переменных равен $2^n \cdot n!$. Исходя из этого, оценим возможность применения тривиальных алгоритмов, перечисляющих все элементы группы. Время проверки одного элемента составляет $\tau(n) = 10^{-9} \cdot 2^n$,

[1] Geissinger L., Kinch D. Representations of the Hyperoctahedral Groups, Journal of algebra, University of North Carolina, 1978, vol. 53, p. 1–20.

[2] Gates W. H., Papadimitriou C. H. Bounds for sorting by prefix-reversal, Discrete Mathematics, 1979, vol. 27, p. 47–57.

[3] Hannenhalli S., Pevzner P.A. Transforming cabbage into turnip (polynomial algorithm for sorting signed permutation by reversal), J. ACM, 1999, vol. 46, no. 1, p. 1–27.

[4] Тарасов А. В. Некоторые свойства групп инерции булевых биюнктивных функций и индуктивный метод генерации таких функций, Дискретная математика, 2002, т. 14, № 2, с. 33–47.

[5] Винокуров С. Ф., Казимиров А. С. Перечисление операторных классов булевых функций, Известия Иркутского государственного университета. Серия: Математика, 2009, т. 2, № 2, с. 40–55.

[6] Погорелов Б. А., Пудовкина М. А. Свойства графов орбиталов налгрупп группы Джевонса, Математические вопросы криптографии, 2010, т. 1, № 1, с. 55–83.

[7] Алексеев Е. К., Карелина Е. К. Классификация корреляционно-иммунных и минимальных корреляционно-иммунных булевых функций от 4 и 5 переменных, Дискретная математика, 2015, т. 27, № 1, с. 22–33.

где 2^n – число значений БФ и 10^{-9} с – примерное время вычисления одного значения на ЭВМ (1 ГГц). Тогда полное время работы алгоритма составит $Time(n) = \tau(n) \cdot 2^n \cdot n!$, например, для $n = 19$ получится более триллиона лет. В работах С. Голомба[8] и Э. А. Якубайтиса[9] предлагались решения указанных проблем, однако сложность предложенных решений сопоставима со сложностью тривиального алгоритма. По этой причине джевонс-эквивалентные преобразования используются в качестве криптографических примитивов в алгоритмах шифрования, основанных на управляемых операциях, где элемент группы является ключом шифрования, а БФ – исходными данными и шифротекстом.

Решения указанных задач позволят в перспективе разработать алгоритмы анализа данных, эквивалентных относительно других групп. Задачи, решаемые такими алгоритмами, появляются естественным образом в прикладных областях. Показательным примером является задача решения уравнения действия аддитивной группы кольца вычетов над данными, эквивалентными БФ, при приёме спутникового сигнала ГЛОНАСС. Наибольший интерес представляют разработка и исследование моделей и алгоритмов обработки информации, основывающихся на операциях над классами джевонс-эквивалентных данных. Исследования в этом направлении являются теоретической основой для разработки качественно новых алгоритмов сжатия данных. Исследования операций над джевонс-эквивалентными данными (над джевонс-эквивалентными БФ) позволят также разработать более точные методы распознавания образов. Отдельные направления работ позволят создать методы помехоустойчивого кодирования при условии невозможности добавления избыточности комбинаторными методами (задача восстановления повреждённого спутникового сигнала).

Объект исследования. Джевонс-эквивалентные данные.

Предмет исследования. Анализ джевонс-эквивалентности данных.

Целью диссертационной работы является создание эффективных алгоритмов анализа джевонс-эквивалентности данных и вычисление действующих элементов группы Джевонса.

Поставленная цель достигается путем решения следующих **задач**:

а) найти эффективные представления группы Джевонса, необходимые для задания её действия на множествах БВ и БФ;

б) исследовать свойства действия группы Джевонса над БВ и БФ;

в) создать алгоритмы решения уравнения действия элемента группы Джевонса над БФ относительно неизвестного действующего элемента;

г) оценить эффективность предложенных алгоритмов и возможность их применения в прикладных задачах.

[8] Golomb S. W. On classification of Boolean functions, IRE, Trans.circuit theory. Spec.Suppl., 1959, № 6, p. 176–186.

[9] Якубайтис Э. А. Субклассы и классы булевых функций, Автоматика и вычислительная техника, Рига: Зинатне, 1974, № 1, с. 1–8.

Соответствие диссертации паспорту специальности. Диссертационная работа соответствует области исследований специальности 05.13.17 – Теоретические основы информатики по п. 5 «Разработка и исследование моделей и алгоритмов анализа данных» и п. 14 «Разработка теоретических основ создания программных систем для новых информационных технологий».

Методы исследования. Основные результаты получены на основе методов теории информации, теории групп, комбинаторного анализа, дискретной математики и высокопроизводительных вычислений.

Научная новизна:

а) найдены два эффективных представления группы Джевонса для задания действия над БВ и БФ, которые позволяют снижать трудозатраты при разработке моделей программных систем, основанных на этом действии;

б) исследованы действия элемента группы Джевонса над БФ, и в результате найдены новые частотные свойства этих действий. Такие свойства позволяют разрабатывать и исследовать алгоритмы анализа данных, основывающиеся на их частотных (энтропийных) характеристиках;

в) найдено новое каноническое представление элемента группы Джевонса, и на его основе создан эффективный алгоритм решения уравнения действия такого элемента над БФ. Он позволяет решить проблему поиска элементов группы, связывающих джевонс-эквивалентные данные;

г) введено новое понятие эквиморфизма групп, доказано эквиморфное вложение группы Джевонса в симметрическую группу степени 2^n. На его основе разработан эквиморфный вычислитель, являющийся моделью архитектуры процессора, на котором могут создаваться новые программные системы обработки данных. Он включает в себя эффективные алгоритмы вычисления действия элемента группы Джевонса над БФ.

Теоретическая и практическая значимость работы. Работа носит теоретический характер. Результаты могут быть использованы как непосредственно для определения джевонс-эквивалентности данных, так и для разработки и исследования частотных моделей и алгоритмов обработки информации. Отдельные выводы могут быть использованы для анализа безопасности некоторых криптографических алгоритмов и для генерации специальных данных с одинаковыми частотными (энтропийными) характеристиками **во всех их допустимых алфавитах**.

Положения, выносимые на защиту диссертационной работы. На защиту выносятся следующие основные результаты:

а) представления группы Джевонса для действия над БВ и БФ;

б) частотные свойства действия элемента группы Джевонса над БФ и метод формирования БВ с одинаковыми частотными (энтропийными) характеристиками **во всех допустимых** для них алфавитах;

в) модель канонического представления элемента группы Джевонса и эффективный алгоритм решения уравнения действия её элемента над БФ;

г) модель эквиморфного вычислителя и его эффективные алгоритмы.

Достоверность результатов работы подтверждается математическими доказательствами основных положений. Эффективность предлагаемых алгоритмов подтверждается результатами, полученными на основе методов спектрального анализа БФ и вычислительных экспериментов.

Апробация результатов работы. Результаты диссертационной работы докладывались автором на следующих международных конференциях: Международная конференция «Алгебра и логика: теория и приложения» (Красноярск, 2013 г.); XVIII, XX Международные конференции «Решетнёвские чтения» (Красноярск, 2014, 2016 гг.); Международные конференции Мальцевские чтения 2014, 2016 (Новосибирск, 2014, 2016 гг.); IX Международная конференция «Дискретные модели в теории управляющих систем» (Москва, 2015 г.).

Результаты работы неоднократно обсуждались на семинарах в Сибирском государственном аэрокосмическом университете, Сибирском федеральном университете, а также в Институте математики им. С.Л. Соболева СО РАН.

Публикации. По результатам диссертационного исследования опубликовано 14 печатных работ, из которых 4 изданы в журналах, рекомендованных ВАК, 8 – в тезисах и трудах конференций, и 2 свидетельства о регистрации программы, зарегистрированных в Российском реестре программ для ЭВМ. 6 из 14 печатных работ опубликованы в неразделимом соавторстве с научным руководителем А. А. Кузнецовым. Тезисы и труды конференций не приводятся в связи с наличием перечня международных конференций, на которых были представлены основные результаты исследования.

Структура работы. Диссертационная работа изложена на 119 листах и состоит из введения, четырёх глав, заключения, списка литературы, списка сокращений и условных обозначений и четырёх приложений.

Содержание диссертации

Во **введении** обоснована актуальность темы диссертационной работы, определены цель и задачи исследования, указаны применяемые в работе методы, представлены основные результаты.

В **первой главе** решается задача выбора конструктивного представления группы Джевонса, необходимого для действий над БВ и БФ, и вводятся основные обозначения. Пусть $n, k = 2^n, i, j$ – целые неотрицательные числа. $E = \{0, 1\}$ – бинарное множество, E^n – множество БВ длины n. Для координат БВ используется нотация L2R (left to right), т.е. бо́льшие координаты находятся левее. $E_n = (E^n, \oplus)$ – группа линейных сдвигов, S_n – симметрическая группа, действующая на множестве $[0; n-1]$. Определим действие S_n на E^n как:

$$x'_{\pi(i)} = x_i; \qquad (1.1^{\mathrm{A}})$$

$$x'_i = x_{\pi(i)}. \tag{1.1^Б}$$

В зависимости от выбора типа действия композиция нескольких действий одного типа будет рассчитываться по-разному.

Лемма 1.1^А (*О действии типа А*). *Если подстановки $\pi, \rho \in S_n$ действуют последовательно (π и затем ρ) над элементом $x \in E^n$, то результат действия эквивалентен действию подстановки-произведения $\pi\rho$, или в символьном виде $(x^\pi)^\rho = x^{(\pi\rho)}$.*

Лемма 1.1^Б (*О действии типа Б*). *Если подстановки $\pi, \rho \in S_n$ действуют последовательно (π и затем ρ) над элементом $x \in E^n$, то результат действия эквивалентен действию подстановки-произведения $\rho\pi$, или в символьном виде $(x^\pi)^\rho = x^{(\rho\pi)}$.*

Действие S_n задаётся также на E_n и реализует автоморфизм.

Теорема 1.1 (*Об автоморфизме*). *Отображение $E_n^\pi \to E_n, \forall \pi \in S_n$ есть автоморфизм независимо от типа действия А или Б.*

Представления группы Джевонса определяются гомоморфизмами внешнего полупрямого произведения, которые различны и могут давать неизоморфные группы (например тождественный гомоморфизм). Для выбора гомоморфизма построим группу, которая индуцирует эквивалентное действие на множестве БФ и раскладывается во внутреннее полупрямое произведение. Вложим в неё группу Джевонса и из внутреннего автоморфизма определим внешний.

$$\forall x \in E^n, \exists! x \in \mathbb{Z}_+ \colon x = \sum_{i=n-1}^{0} x_i \cdot 2^i. \tag{1.2}$$

Отображение (1.2) является биекцией и $0 \leqslant x < k$. Порядок индексов является L2R. Определим действие элементов E_n и S_n над элементами $\mathbb{Z}_+$ как:

$$x^z = \sum_{i=n-1}^{0} (x_i \oplus z_i) \cdot 2^i; \tag{1.3}$$

$$x^\pi = \sum_{i=n-1}^{0} x_{\pi^{-1}(i)} \cdot 2^i = \sum_{i=n-1}^{0} x_i \cdot 2^{\pi(i)}; \tag{1.4^А}$$

$$x^\pi = \sum_{i=n-1}^{0} x_{\pi(i)} \cdot 2^i = \sum_{i=n-1}^{0} x_i \cdot 2^{\pi^{-1}(i)}. \tag{1.4^Б}$$

Откуда подстановки из S_k будут ($0 \leqslant j < n$):

$$\varphi_z = \begin{pmatrix} 2^n - 1 & \dots & j & \dots & 0 \\ \sum_{i=n-1}^{0} \overline{z_i} \cdot 2^i & \dots & \sum_{i=n-1}^{0} (j_i \oplus z_i) \cdot 2^i & \dots & \sum_{i=n-1}^{0} z_i \cdot 2^i \end{pmatrix}; \tag{1.5}$$

$$\varphi_\pi = \begin{pmatrix} 2^n - 1 & \dots & j & \dots & 0 \\ 2^n - 1 & \dots & \sum_{i=n-1}^{0} j_i \cdot 2^{\pi(i)} & \dots & 0 \end{pmatrix}; \tag{1.6^А}$$

$$\varphi_\pi = \begin{pmatrix} 2^n - 1 & \dots & j & \dots & 0 \\ 2^n - 1 & \dots & \sum_{i=n-1}^{0} j_{\pi(i)} \cdot 2^i & \dots & 0 \end{pmatrix}. \tag{1.6^Б}$$

Подстановки (1.5) и (1.6^А) или (1.6^Б) действуют над БВ длины k, которые эквивалентны БФ. Пусть B_n и T_n – множества подстановок (1.5) и (1.6^А) или (1.6^Б) соответственно. Они подгруппы в S_k и являются вложениями E_n и S_n.

Теорема 1.2 (*Об изоморфизме* B_n). *Отображение $\varphi_z \colon E_n \to B_n, z \in E_n, \varphi_z \in B_n \colon \varphi_z(j) = j^z, 0 \leqslant j < k$ есть изоморфизм.*

Теорема 1.3 (*Об изоморфизме* T_n). *Отображение $\varphi_\pi \colon S_n \to T_n, \pi \in S_n, \varphi_\pi \in T_n \colon \varphi_\pi(j) = j^\pi, 0 \leqslant j < k$ есть изоморфизм для типа действия А и* *антиизоморфизм для типа действия Б.*

Теоремы 1.2 и 1.3 позволяют вложить D_n в S_k и $D_n \to \beta_n < S_k\colon \beta_n = B_n \leftthreetimes T_n$.

Теорема 1.4 (***Об изоморфизме*** β_n). *Множество $\beta_n = B_n \cdot T_n$, образованное как теоретико-множественное произведение подгрупп B_n и T_n, есть группа, которая является внутренним полупрямым произведением $B_n \leftthreetimes T_n$.*

Откуда получим два представления группы Джевонса $(z_0\pi_0), (z_1\pi_1) \in D_n$:

$$(z_0\pi_0)(z_1\pi_1) = \left(z_0 z_1^{\pi_0^{-1}} \pi_0\pi_1\right); \tag{1.7^А}$$

$$(z_0\pi_0)(z_1\pi_1) = \left(z_0 z_1^{\pi_0^{-1}} \pi_1\pi_0\right). \tag{1.7^Б}$$

Основные результаты первой главы опубликованы в [1].

Во **второй главе** определяются и исследуются действия группы Джевонса и её подгрупп над БВ и БФ. Определим действие $(z\pi) \in D_n$ над БВ:

$$x^{(z\pi)} = (x^z)^\pi. \tag{2.1}$$

Лемма 2.1 (***О композиции действий над бинарным вектором***). *Последовательное действие двух элементов группы Джевонса $(z_0\pi_0), (z_1\pi_1) \in D_n$ на бинарный вектор $x \in E^n$ эквивалентно одному действию произведения этих элементов в исходном порядке по внутригрупповой операции, или в символьном виде $\left(x^{(z_0\pi_0)}\right)^{(z_1\pi_1)} = x^{(z_0\pi_0)(z_1\pi_1)}$.*

Пусть X (множество значений аргументов) и Y (множество значений) некоторых функций f, g и пусть g – результат преобразования аргумента f:

$$f, g\colon X \to Y, \mu\colon X \to X, g = f^\mu\colon g(x) = f(x^\mu), \forall x \in X. \tag{2.2}$$

Лемма 2.2 (***О композиции действий над булевой функцией***). *Если над функцией $f\colon X \to Y$ проводятся подряд действия через аргумент $\mu, \nu\colon X \to X$, то результат их эквивалентен функции, аргумент которой преобразуется этими же действиями в обратном порядке, или в символьном виде $(f^\mu)^\nu = f((x^\nu)^\mu)$.*

БФ n аргументов $f(x_{n-1}, \ldots, x_i, \ldots, x_0)$ – отображение $E^n \to E$. Действие $(z\pi) \in D_n$ над ней согласно (2.2) и их композицию по лемме 2.2 определим как:

$$f^{(z\pi)} = f\left(x^{(z\pi)}\right) = f((x^z)^\pi) = \left(f^{(e_E\pi)}\right)^{(ze_S)}, \left(f^{(z_0\pi_0)}\right)^{(z_1\pi_1)} = f^{(z_1\pi_1)(z_0\pi_0)}. \tag{2.3}$$

Поставим в соответствие булевой функции f БВ y по правилу:

$$y = \{f(11\ldots11), f(11\ldots10), \ldots, f(00\ldots01), f(00\ldots00)\}, y_j = f(j). \tag{2.4}$$

Тогда над (2.4) действует группа β_n, – и это эквивалентно действию группы Джевонса над самой БФ. Определим новое понятие – «эквимирфизм групп».

Определение 2.1. *Две группы G, G', действующие на некотором множестве M, будем называть **эквиморфными**, если существует биекция $\varphi\colon G \to G'$, такая, что для любых $a, b \in G$ и $m \in M$:*

$$(m^a)^b = \left(m^{\varphi(a)}\right)^{\varphi(b)}. \tag{2.5}$$

При этом отображение φ будем называть эквиморфизмом.

Теорема 2.1 (***Об эквиморфизме групп Джевонса и*** β_n). *Группа Джевонса D_n эквиморфна группе β_n по действию над БФ по типу Б и экви-*

морфна в обратные (отрицательные) элементы β_n по действию над БФ по типу А, и эквиморфизмами будут композиции отображений (1.5), ($1.6^{\text{А}}$) и ($1.6^{\text{Б}}$).

Для типа А $(z\pi) \to \varphi^{-1}_{(z\pi)} = \varphi^{-1}_{\pi}\varphi^{-1}_{z}$ и для типа Б $(z\pi) \to \varphi_{(z\pi)} = \varphi_z\varphi_\pi$.

БВ могут рассматриваться как данные над различными алфавитами.

Определение 2.2. *Алфавит A_i – множество E^{2^i}.*

Определение 2.3. *Символ алфавита – элемент алфавита A_i.*

БВ y_f может быть разбит на символы в любом из алфавитов A_i, где i пробегает все значения $[0;n]$. При этом разбиение начинается с первого значения вектора y_f и символы не перекрываются. Длина БВ y_f (число символов) в каждом из алфавитов рассчитывается как $l_i = 2^{n-i}$.

Определение 2.4. *Частота символа – число случаев встречи заданного символа из алфавита A_i в векторе y_f.*

Определение 2.5. *Частотное распределение БВ y_f (спектр) над алфавитом A_i – отношение $Q_i(y_f) \subset A_i \times [0;k]$.*

Определение 2.6. *Спектральное распределение БВ y_f над алфавитом A_i – отношение $R_i(y_f) \subset [0;k] \times [0;k]$, элементы которого показывают, как часто повторяются частоты в векторе y_f.*

Пусть q_v – относительная частота символа v данных, тогда энтропия:

$$H(y) = -\sum_v q_v \log_2 q_v = -\sum_v \log_2 q_v^{q_v}. \tag{2.6}$$

Опираясь на теорему 2.1, удалось найти следующие частотные свойства действия группы Джевонса над БФ. Рассмотрим теоремы, их доказывающие.

Определим $c_{n-1}, \ldots, c_0 \in E^n$, причём каждый c_i имеет на позиции i значение 1, а на остальных – 0.

Теорема 2.2 (***Об инвариантности спектров при действии*** E_n)**.** *Если элемент c_i действует на БФ $f \in B(n)$, то частотные распределения БВ y_f инвариантны относительно этого действия для алфавитов $A_{i'} : i' \leqslant i, i \in [0; n-1]$, а спектральные распределения БВ y_f инвариантны относительно этого же действия для алфавитов A_i.*

Следствие 2.2.1. *Энтропия БВ y_f инвариантна во всех алфавитах $A_i, i \in [0; n-1]$ относительно действия любого элемента E_n над БФ f.*

Теорема 2.3 (***Об инвариантности спектров при действии*** S_n)**.** *Если транспозиция $(i,j) \in S_n, i,j \in [0; n-1] : i < j$ действует на БФ $f \in B(n)$, то частотные распределения БВ y_f инвариантны относительно этого действия для алфавитов $A_{i'} : i' \leqslant i$, а спектральные распределения БВ y_f инвариантны относительно этого же действия для алфавитов $A_{i'}$ и $A_{j'} : j' > j$.*

Следствие 2.3.1. *Энтропия БВ y_f инвариантна для алфавитов индексов до i включительно и больше j относительно действия транспозиции $(i,j), i,j \in [0; n-1] : i < j$ группы S_n над БФ f.*

Согласно теореме 2.2, энтропия сохраняется **во всех допустимых алфа-**

витах данных. Само действие E_n является методом генерации таких данных, и его эффективность можно оценить через порядок подгруппы инерции.

Определение 2.7. *Подгруппой инерции БФ f в группе G называть подмножество, элементы которого действуют над f тривиально, или в символьном виде $J_G(f) = \{g \in G \mid f^g = f\}$.*

Число различных БФ равно $[G : J_G(f)]$, и согласно теории Пойа верно:

$$count_{E_n} \approx 2^n. \tag{2.7}$$

Основные результаты второй главы опубликованы в [2, 3].

В **третьей главе** рассматриваются алгоритмы эффективного решения уравнения действия элемента группы Джевонса над БФ вида:

$$f^{(z\pi)} = g, \tag{3.1}$$

где $f, g \in B(n)$ – исходная и результирующая булевы функции соответственно и $(z\pi)$ – неизвестный действующий элемент.

Алгоритмы основываются на следующих опорных утверждениях.

Лемма 3.1 (*О монотонном представлении подстановки*). *Пусть k есть количество независимых циклов, включая циклы длины 1, нетривиальной подстановки π группы S_n степени n. Тогда она может быть единственным образом представлена как произведение из $n-k$ транспозиций:*

$$\pi = \left(0, \pi_0^{-1}(0)\right) \cdots \left(i_0, \pi_{i_0}^{-1}(i_0)\right) \cdots \left(i, \pi_i^{-1}(i)\right) \cdots \left(i_1, \pi_{i_1}^{-1}(i_1)\right) \cdots \left(n-2, \pi_{n-2}^{-1}(n-2)\right), \tag{3.2}$$

причём $0 \leqslant i_0 < i < i_1 < n-1$. В произведение включаются только транспозиции, соответствующие точкам i: $\pi(i) \neq i$ и $i \leqslant \pi_i^{-1}(i)$. Промежуточные подстановки вычисляются рекурсивно как $\pi_{i+1} = \left(i, \pi_i^{-1}(i)\right)\pi_i$, при этом $\pi_0 = \pi$.

Пусть $b_{n-1}, \ldots, b_i, \ldots, b_0, z \in E_n$, причём $b_i = \{0, \ldots, 0, \ldots, z_i, \ldots, 0, \ldots, 0\}$, т.е. содержит на позиции i значение координаты i БВ z, а на остальных – нули.

Теорема 3.1 (*О каноническом представлении элемента группы Джевонса*). *Любой элемент группы Джевонса $(z\pi) \in D_n$ представим единственным образом в виде произведения:*

$$(b_{n-1}(n-1, j_{n-1})) \cdots (b_{i_1}(i_1, j_{i_1})) \cdots (b_i(i, j_i)) \cdots (b_{i_0}(i_0, j_{i_0})) \cdots (b_0(0, j_0)), \tag{3.3}$$

где $0 \leqslant i_0 < i < i_1 \leqslant n-1$. Элементы симметрической группы (i, j_j): $i < j_i$ имеют порядок не более 2 и, в случае транспозиции, соответствуют представлению по лемме 3.1 (для типа А будет π^{-1} и для типа Б – π).

Запишем уравнение (3.1), опираясь на (3.3), следующим образом:

$$\left(\left(\left(\left(f^{(b_0(0,j_0))}\right)^{\cdots\left(b_{i_0}\left(i_0,j_{i_0}\right)\right)}\right)^{\cdots(b_i(i,j_i))}\right)^{\cdots\left(b_{i_1}\left(i_1,j_{i_1}\right)\right)}\right)^{\cdots(b_{n-1}(n-1,j_{n-1}))} = g. \tag{3.4}$$

На основе формы уравнения (3.4) можно построить **Алгоритм 3.1 (*О вычислении нулевого действия*)**. Он вычисляет все решения уравнения (3.1) и основывается на последовательном вычислении множителей, в них входящих, согласно представлению (3.3). Рассмотрим следующую теорему.

Теорема 3.2 (*О вычислении нулевого действия*). *Множество H_n*

совпадает с множеством всех решений уравнения $f^{(z\pi)}$ *и может быть вычислено как* $H'_{i+1}{=}\left[\bigcup\limits_{j_i=i}^{j_i<n}(e_E(i,j_i))H_i\right]\bigcup\left[\bigcup\limits_{j_i=i}^{j_i<n}(c_i(i,j_i))H_i\right]$ *за* n *шагов последовательно для* $0{\leqslant}i{<}n$ *и* $H_{i+1}{=}\left\{h\in H'_{i+1}\mid Q_{i+1}\left(f^h\right){=}Q_{i+1}(g)\right\}$*, начиная с* $H_0{=}\{(e_E e_S)\}$*.*

Откуда число действий над БФ в алгоритме 3.1 составит:

$$d=\sum\nolimits_{i=0}^{n-1} r_i\cdot 2\cdot(n-i), r_i=|H_i|\,. \tag{3.5}$$

Алгоритм 3.1 вычисляет действия множителей представления (3.3). Можно существенно повысить эффективность их вычисления за счёт архитектуры процессоров и выводов теоремы 2.1. Эти множители могут содержать отрицание или транспозицию или отрицание и транспозицию вместе. Для вычисления действий таких элементов разработаны **Алгоритм 3.2 (*Об эквиморфном вычислении действия* E_n)**, **Алгоритм 3.3 (*Об эквиморфном вычислении действия* S_n)** и **Алгоритм 3.4 (*Об эквиморфном вычислении действия* D_n)**. Предлагается модель эквиморфного вычислителя, реализующего примитивные операции над БВ: логические умножение и сложение, логические левый и правый сдвиги на число и присвоение.

Определение 3.1. *Слово – бинарный вектор длины* $2^{n'}$*, над которым вычислителем выполняются примитивные операции, где* n' *– его степень.*

Для предлагаемых далее алгоритмов выполняется разбиение БВ на слова числом $w=2^{n-n'}$ и производится их вычисление так, что каждое слово и его значения обрабатываются только один раз. Далее рассмотрим теоремы, доказывающие их корректность и оценивающие их сложность.

Теорема 3.3 (*Об эквиморфном вычислении действия* E_n). *Вычисление БВ* $y\in E^k$*, эквивалентного БФ* $f\in B(n)$*, по алгоритму 3.2 равносильно действию эквиморфизма* φ_{c_i} *(по формуле (1.5)) над ним. Число операций эквиморфного вычислителя для* $i<n'$ *составит* $2{\cdot}w$ *сдвигов,* $2{\cdot}w$ *умножений,* w *сложений,* w *присвоений, а для* $i\geqslant n'$ *–* w *присвоений.*

Теорема 3.4 (*Об эквиморфном вычислении действия* S_n). *Вычисление БВ* $y{\in}E^k$*, эквивалентного БФ* $f{\in}B(n)$*, по алгоритму 3.3 равносильно действию эквиморфизма* $\varphi_{(i,j)}$ *(по формуле* $(1.6^{\text{А}})$ *или* $(1.6^{\text{Б}})$*) над ним. Число операций эквиморфного вычислителя для* $j{<}n'$ *составит* $2{\cdot}w$ *сдвигов,* $3{\cdot}w$ *умножений,* $2{\cdot}w$ *сложений,* w *присвоений, а для* $j{\geqslant}n'$ *при* $i{<}n'$ *–* w *сдвигов,* $2{\cdot}w$ *умножений,* w *сложений,* w *присвоений, и при* $i{\geqslant}n'$ *–* w *присвоений.*

Теорема 3.5 (*Об эквиморфном вычислении действия* D_n). *Вычисление БВ* $y\in E^k$*, эквивалентного БФ* $f\in B(n)$*, по алгоритму 3.4 равносильно действию эквиморфизма* $\varphi_{(c_i(i,j))}$ *над ним, получаемого по теореме 2.1. Число операций эквиморфного вычислителя для* $j{<}n'$ *составит* $4{\cdot}w$ *сдвигов,* $4{\cdot}w$ *умножений,* $3{\cdot}w$ *сложений,* w *присвоений, а для* $j{\geqslant}n'$ *при* $i{<}n'$ *–* w *сдвигов,* $2{\cdot}w$ *умножений,* w *сложений,* w *присвоений, и при* $i{\geqslant}n'$ *–* w *присвоений.*

Основные результаты третьей главы опубликованы в [4, 5].

В **четвёртой главе** предлагается оценка сложности алгоритма 3.1 через минимальное и максимальное число действий по формуле (3.5):

$$d_{min} = \sum_{i=0}^{n-1} 2 \cdot (n-i) = 2 \cdot n \cdot \frac{(n-0)+(n-(n-1))}{2} = n^2 + n; \quad (4.1)$$

$$d_{max} = \sum_{i=0}^{n-1} 2^i \cdot \frac{n!}{(n-i)!} \cdot 2 \cdot (n-i). \quad (4.2)$$

Для практического использования алгоритма 3.1 этого недостаточно, т.к. опыт показывает, что в подавляющем большинстве случаев число действий $O(n^2)$. Выполнена эмпирическая оценка этого числа для множества конкретных уравнений, отражающих реальные данные. Для более чем $2^{64} \approx 10^{20}$ уравнений (3.1) были исследованы причины увеличения числа действий выше (4.1) и сформулированы предложения о реальной сложности алгоритма 3.1 и о возможности его применения для решения прикладных задач. Исследование проводилось с применением инженерно-технических решений, основывающихся на специально разработанной библиотеке ***domain object processor*** (или **dop**).

Для БФ с нетривиальной $J_{D_n}(f)$ появление числа действий при решении уравнения (3.1) больше (4.1) следует из доказательства теоремы 3.2, но при этом таких БФ подавляющее меньшинство. Поэтому исследование включает в себя: анализ тривиальности $J_{D_n}(f)$, спектральный анализ всех БФ с тривиальной $J_{D_n}(f)$ для $n = 1, 2, 3, 4, 5$ и статистический анализ числа действий некоторого количества (миллионов) уравнений (3.1) для $5 < n < 24$. Для формирования статистики случайные БВ распределены равномерно, потому что при обработке реальных данных, в общем случае, будет такое же распределение. В табл. 1 приведен расчёт числа действий по (3.5) всех БФ для $n = 4, 5$ с тривиальной $J_{D_n}(f)$ и оценка эффективности, т.е. отношение сложности предлагаемого алгоритма к тривиальному. В табл. 2 приводятся результаты численного эксперимента миллионов уравнений для $5 < n < 24$. Теоретическая эффективность алгоритмов 3.2, 3.3 и 3.4 минимум в $2^{n'}$ раз, при этом эксперимент показал существенно бóльшую эффективность (для $n' = 5$ более чем в 750 раз).

Таблица 1. Эффективность для $n = 4, 5$

n	4		5	
Характеристика	Число действий	Эффективность	Число действий	Эффективность
Лучший случай	20	94,791 7 %	42	99,218 8 %
Средний случай	27,898 3	92,734 8 %	42,273 0	98,899 1 %
Худший случай	42	89,062 5 %	184	95,208 3 %
Тривиальный алгоритм	384	–	3 840	–

Таблица 2. Число действий для $5 < n < 24$ (10^6 экспериментов)

n	6	7	8	9	10	11	12	13	14
n^2+n	42	56	72	90	110	132	156	182	210
Эксперимент	46,028	58.918	74,106	91,469	111,037	132,708	156,479	182,338	210,240

n	15	16	17	18	19	20	21	22	23
n^2+n	240	272	306	342	380	420	462	506	552
Эксперимент	240,169	272,118	306,078	342,066	380,049	420,034	462,019	506,020	552,005

Основные результаты четвёртой главы опубликованы в [6].

В **четырёх приложениях** приведены статистика и обзор классов БФ для $n = 1, 2, 3, 4$ относительно E_n, S_n и D_n, а также результаты спектрального анализа и численных экспериментов.

В **заключении** приведены выводы работы и сформулированы основные результаты.

Основные результаты и выводы

Найдены два два типа представления группы Джевонса: А для действия над БВ и Б – над БФ (теорема 1.4). Выбор типа действия, в зависимости от исследуемого объекта при разработке модели программной системы, существенно снижает трудозатраты за счёт упрощения модели и этапа её проектирования.

Найдены частотные свойства действия элемента группы Джевонса, которые заключаются в инвариантности частотных (энтропийных) характеристик в заданных действующим элементом алфавитах (теоремы 2.2 и 2.3). Это влияние на энтропию позволяет разрабатывать алгоритмы генерации данных для анализа алгоритмов их преобразования. Отдельно стоит подчеркнуть действия E_n, т.к. они сохраняют энтропию **во всех допустимых алфавитах** данных.

Предложена модель канонического представления элемента группы Джевонса (теорема 3.1) и создан новый эффективный алгоритм решения уравнения действия элемента группы Джевонса над БФ (теорема 3.2). Он эффективен при решении уравнений, включающих БФ с тривиальной подгруппой инерции в группе Джевонсна. Таких БФ подавляющее большинство, поэтому показана возможность использования предложенного алгоритма для решения практических задач анализа джевонс-эквивалентности данных. Опираясь на полученные результаты, важно отметить, что появляются сомнения в применении криптографических примитивов в алгоритмах шифрования, основанных на управляемых операциях, где элемент группы Джевонса является ключом шифрования, а БФ – исходными данными и шифротекстом.

Доказано эквиморфное вложение группы Джевонса в симметрическую группу степени 2^n (теорема 2.1). Разработаны эквиморфный вычислитель и его алгоритмы работы (теоремы 3.3, 3.4 и 3.5), позволяющие ещё больше повысить эффективность анализа джевонса-эквивалентности данных. Практическая проверка вычислителя подтверждает возможность интеграции предложенных алгоритмов в программные системы обработки данных.

Полученные результаты позволяют вести исследования в данной и смежных предметных областях в направлениях:

– создание расширенного алгоритма анализа джевонс-эквивалентности данных, позволяющего вычислять порождающие подгруппы инерции БФ в группе Джевонса и находить решения уравнения для произвольных БФ. Разработка расширенного алгоритма включает в себя отдельное исследование нерешённых теоретических и практических задач, таких как определение диаметра графа Кэли для группы Джевонса, заданной множителями канонического представления;

– создание алгоритмов анализа данных, эквивалентных относительно других групп. Задачи, решаемые такими алгоритмами, появляются естественным образом в прикладных областях. Показательным примером является задача решения уравнения действия аддитивной группы кольца вычетов над данными, эквивалентными БФ, при приёме спутникового сигнала ГЛОНАСС;

– разработка и исследование моделей и алгоритмов обработки информации, основывающихся на операциях над классами джевонс-эквивалентных данных. Исследования в этом направлении наиболее интересны, потому что являются теоретической основой для разработки качественно новых алгоритмов сжатия данных. Исследования операций над джевонс-эквивалентными данными (над джевонс-эквивалентными БФ) позволят также разработать более точные методы распознавания образов. Отдельные направления работ позволят создать методы помехоустойчивого кодирования при условии невозможности добавления избыточности комбинаторными методами (задача восстановления повреждённого спутникового сигнала).

Публикации по теме диссертации

В изданиях, рекомендованных ВАК:

1. Кукарцев, А. М. О конструктивном представлении группы Джевонса для инженерно-технических решений обработки информации / А. М. Кукарцев, А. А. Кузнецов // Программная инженерия. – М., 2015. – № 11. – С. 25–33.

2. Кукарцев, А. М. О действиях группы Джевонса на множествах бинарных векторов и булевых функций для инженерно-технических решений обработки информации / А. М. Кукарцев, А. А. Кузнецов // Программная инженерия. – М., 2016. – Т. 7. – № 1. – С. 29–36.

3. Кукарцев, А. М. О частотных свойствах действий группы Джевонса на булевых функциях / А. М. Кукарцев // Программная инженерия. – М., 2016. – Т. 7. – № 11. – С. 515–521.

4. Кукарцев, А. М. Об эффективном алгоритме решения уравнения действия группы Джевонса над булевыми функциями / А. М. Кукарцев, А. А. Кузнецов // Программная инженерия. – М., 2017. – Т. 8. – № 2. – С. 76–87.

Свидетельства о государственной регистрации программ для ЭВМ:

5. Кукарцев, А. М. Библиотека domain object processor (dop) / А. М. Кукарцев. – Свидетельство о государственной регистрации программы для ЭВМ № 2016615233 от 18.05.2016 г.

6. Кукарцев, А. М. Программный комплекс спектрального анализа булевых функций SpectrumAnalyzer / А. М. Кукарцев. – Свидетельство о государственной регистрации программы для ЭВМ № 2016615313 от 20.05.2016 г.

Автор выражает свою благодарность Богу, Кукарцевой Татьяне Дмитриевне и Кукарцеву Михаилу Владимировичу – маме и папе, научному руководителю Кузнецову Александру Алексеевичу за плодотворное руководство, стойкость, смелость, терпение и поддержку на всех этапах выполнения работы, Лубкину Ивану Александровичу за помощь в развитии основных идей в рассматриваемой предметной области, Цареву Сергею Петровичу и коллективу журнала «Программная инженерия» за ценные замечания и рекомендации при подготовке текстов статей, Глухову Михаилу Михайловичу и Белякову Геннадию Павловичу, а также Агибалову Геннадию Петровичу за помощь в поиске материалов, Созутову Анатолию Ильичу и Шлепкину Анатолию Константиновичу за консультации и помощь в представлении основных результатов, Попову Алексею Михайловичу, Сафонову Константину Владимировичу, Быковой Валентине Владимировне, Покидышевой Людмиле Ивановне и Грудиновой Татьяне Абрамовне за неоценимую помощь в подготовке работы. Отдельно автор выражает благодарность за помощь в научном исследовании коллегам из Института математики им. С.Л. Соболева СО РАН, и особенно Токаревой Наталье Николаевне и Коломеецу Николаю Александровичу, коллегам из института Информатики и телекоммуникаций Сибирского государственного аэрокосмического университета имени академика М.Ф. Решетнёва, и особенно Колесникову Сергею Геннадьевичу, коллегам из института Инженерной физики и радиоэлектроники Сибирского федерального университета, и особенно Саломатову Юрию Петровичу.

Кукарцев Анатолий Михайлович

Эффективные алгоритмы анализа
джевонс-эквивалентности данных

Автореферат диссертации
на соискание учёной степени
кандидата физико-математических наук

Подписано в печать «___»__________ 2017 г. Формат 60×84 1/16.
Бумага офисная. Печать плоская. Усл. печ. л. 1,0. Уч.-изд. л. 1,0.
Тираж 100 экз. Заказ №___.

Отпечатано в отделе копировально-множительной техники СибГАУ
660037, г. Красноярск, пр. им. газ. «Красноярский рабочий», 31.