

Отзыв
официального оппонента

доктора технических наук, доцента, профессора кафедры
Алгебры и математической логики ФГБОУ ВО НГТУ
Чехонадских Александра Васильевича

на диссертацию

«Эффективные алгоритмы анализа джевонс-эквивалентности данных»
на соискание учёной степени кандидата физико-математических наук
по специальности 05.13.17 – Теоретические основы информатики
Кукарцева Анатолия Михайловича

Актуальность темы диссертационной работы. Диссертационное исследование А.М. Кукарцева посвящено теоретико-групповым и алгоритмическим вопросам эффективного представления и обработки данных, представленных в цифровой форме. Действие симметрической группы на булевых векторах и другие близкие вопросы представлений конечных групп привлекали значительно меньшее внимание специалистов по теории групп и их приложений, чем, скажем, представления линейных алгебраических групп над различными полями. Поэтому выбор темы исследования можно оценить не только из-за приложений к теории информации, продемонстрированных в диссертации и имеющих самостоятельное значение, но и как шаг в развитии классической теории симметрических групп и их представлений.

В работе рассматривается один из способов представления данных в виде набора значений (булева вектора) некоторой дискретной (булевой) функции. Над множеством таких данных задаётся отношение эквивалентности посредством действия элементов группы Джевонса D_n . Элемент группы преобразует аргументы булевой функции и, как следствие, содержащиеся в ней данные. Задачи определения эквивалентности двух булевых функций относительно группы Джевонса (и, тем самым, эквивалентности данных) сформулированы более ста лет назад, но не имели нетривиальных решений. А тривиальные решения сводятся к перебору всех элементов группы Джевонса, число которых зависит от размерности задачи как $2^n \cdot n!$ По этой причине указанные задачи отнесены к классу сложно решаемых математических задач. Они естественным образом появляются при реализации цифровых средств обработки информации (в частности, стоит упомянуть известный Гарвардский каталог), а в дискретной математике относятся к области классификации функций.

Соискателем предлагается серия алгоритмов, которые позволяют решать указанные задачи значительно более эффективно, в ряде случаев – за полиномиальное время, что и характеризует актуальность тематики диссертации и ценность ее результатов.

Основная идея диссертационного исследования сводится к тому, что для решения задач представления и сравнения данных применяются методы теории информации и теории групп. Решение строится на двух математических основаниях, непосредственно друг с другом не связанных. Первое заключается в обнаруженных и доказанных автором специфических свойствах действий элементов группы Джевонса на булевы функции, которые заключаются в сохранении частот символов эквивалентных им данных в различных

алфавитах. Второе доказанное автором утверждение сводится к тому, что действующий элемент может быть разделён на части (по аналогии с разрядами десятичных чисел), и каждую часть можно обрабатывать отдельно, опираясь на состав частот символов. (Ближайшей аналогией основной идеи является анализ парольной системы; например, шестизначный цифровой пароль задаёт 1 000 000 комбинаций; если же некоторая система позволяет обрабатывать каждый разряд в отдельности, то число комбинаций уменьшается логарифмически до 60, т.е. 6 разрядов по 10 значений в каждом).

Содержание диссертации. Диссертация изложена на 119 листах и включает в себя введение, четыре главы основного текста, заключение, список литературы, список сокращений и условных обозначений, два справочных и два обязательных приложения.

Во **введении** обоснована актуальность темы и показана степень её разработанности, определены объект и предмет исследования, сформулированы цель и задачи, указаны методы исследования, перечислены новые результаты и охарактеризована структура работы.

В **главе 1** исследуется вопрос выбора представления группы Джевонса, удобного для работы в этой предметной области. Сложность выбора представления заключается в том, что группа Джевонса определяется как внешнее полупрямое произведение групп; и в зависимости от выбора гомоморфизма, входящего в конструкцию произведения, могут быть получены неизоморфные друг другу результаты. В итоге автором предложено два типа представления группы Джевонса: тип А и тип Б. Эти представления сводимы друг к другу (как прямое и обратное), но выбрать только одно не удаётся, потому что тип А более удобен для задания действий над данными, а тип Б – над функциями.

В **главе 2** рассматриваются действия группы Джевонса на множествах данных (булевых векторов) и булевых функций. Вводится ряд понятий, в том числе частотного и спектрального распределений, формулируются и доказываются утверждения, связанные с действием групп сдвигов, подстановок и Джевонса. В этой же главе вводится понятие эквиморфизма групп, то есть эквивалентности групп по действию на множестве. Глава 2 содержит описания и доказательства частотных свойств действия группы Джевонса над булевыми функциями. Заключается глава методом генерации данных с одинаковой энтропией во всех допустимых для них алфавитах.

В **главе 3** рассматривается серия алгоритмов для анализа джевонс-эквивалентности данных. Первый алгоритм предназначен для вычисления действующих элементов группы, связывающих две булевы функции за предположительно полиномиальное время. Вторым, третьим и четвёртым алгоритмы составляют так называемый эквиморфный вычислитель. Он является моделью вычислительного средства, предназначенного для параллельного вычисления результата действия элемента группы Джевонса над булевой функцией. В результате скорость вычисления результата анализа джевонс-эквивалентности двух данных возрастает на несколько порядков. Поэтому основной алгоритм анализа джевонс-эквивалентности данных может применяться к реальным информационным объектам. Предложенные алгоритмы имеют математические доказательства их корректности.

Глава 4 является продолжением главы 3. В ней показывается возможность применения предложенных алгоритмов для анализа джевонс-эквивалентности данных различного размера и статистически доказываются эффективность предложенных решений.

В **заключении** сделан вывод по предложенным алгоритмам и поставлены последующие исследовательские задачи.

Список литературы содержит ссылки на исследования, связанные с группой Джевонса, как отечественных, так и зарубежных научных школ и включает 72 источника.

Справочные приложения содержат сведения о статистике встречаемости джевонс-эквивалентных данных, соответствующих функциям от 1 до 5 аргументов.

Обязательные приложения включают в себя статистические результаты численных экспериментов, доказывающие эффективность полученных решений.

По теме диссертации опубликовано 14 печатных работ, среди которых 4 статьи в изданиях, входящих в перечень ВАК РФ и 2 свидетельства о регистрации программ в Государственном реестре регистрации программ для ЭВМ, а также 8 работ в других изданиях (рецензируемые материалы авторитетных международных конференций).

Текст диссертации удовлетворительно структурирован, и каждая глава заканчивается выводами по ней. В заключении работы приведены общие выводы, соответствующие поставленным во введении цели и задачам. Заявленная тема убедительно раскрыта.

Автором получены следующие **новые научные результаты**.

1. Найдены два эффективных представления группы Джевонса для задания действия над булевыми векторами и булевыми функциями, которые позволяют снизить трудозатраты при разработке моделей программных систем, основанных на этом действии.

2. 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.

3. 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.

4. Введено новое понятие эквиморфизма групп, доказано эквиморфное вложение группы Джевонса в подгруппу симметрической группы 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.

Указанные новые результаты являются преимущественно теоретическими. Основным результатом исследования является эффективный алгоритм анализа джевонс-эквивалентности данных. К имеющим самостоятельную ценность теоретическим результатам можно отнести: доказательство существования монотонного представления подстановки и канонического представления элемента группы Джевонса, а также предложенные конструктивные методы их получения. Вторым из наиболее важных результатов работы является обоснование специальных частотных свойств действий группы Джевонса над булевыми функциями. Эти свойства могут быть различными способами использованы при анализе средств обработки данных. Полученные эффективные представления группы Джевонса для работы в рассматриваемой предметной области, несмотря на строгие доказательства в форме лемм и теорем, являются вспомогательными. Отдельно хочется отметить новое и при этом вполне естественное понятие эквиморфизма групп, которое может быть с успехом использовано в теории представлений групп.

Достоверность полученных результатов. Заявленные в диссертации математические утверждения имеют необходимые доказательства. Для оценки сложности предложенного основного алгоритма анализа джевонс-эквивалентности данных проведены массовые численные эксперименты, результаты которых корректно статистически обработаны. Результаты работы представлялись на авторитетных международных конференциях и обсуждались на семинарах. По материалам работы опубликовано 12 печатных работ, среди которых 4 в статьях рецензируемых изданий, рекомендованных ВАК РФ и 8 работ в рецензируемых трудах международных конференций. Для проведения численных экспериментов разработаны программные комплексы, на которые получены 2 свидетельства о регистрации в Государственном реестре регистрации программ для ЭВМ.

Замечания:

1. Автор применяет в рамках всей работы единый порядок записи, соответствующий *big-endian* (от старшего к младшему). Такой порядок удобен при работе с числами, но автор распространяет её практически на все рассматриваемые им объекты. Для компьютерной обработки информации он эффективен, но в классических учебниках и монографиях рассматриваемые объекты имеют иную форму представления (подстановки, аргументы булевых функций, бинарные вектора). В начале ознакомления с работой это затрудняет чтение и понимание; действия подстановок «типа А» и «типа Б» также естественнее было бы назвать прямым и обратным действием.

2. Определённый на стр. 18 тип действия «Б» и его композиция, рассмотренная в лемме 1.Б на стр. 19, корректны в том смысле, что все последующие выводы формул не содержат противоречий и могут быть использованы для разработки инженерно-технических решений. Однако стоит отметить, что такое определение действия группы на множестве далеко не общеприняты. Заявление автора в том, что невозможно одновременно задать действие и над бинарными векторами, и над булевыми функциями, требует дополнительного разбора.

3. Теоремы 1.4 (стр. 31) и 2.3 (стр. 52) сформулированы не вполне корректно. В первом случае только из доказательства понятно, что понимается под «теоретико-множественным произведением подгрупп». Во втором случае спектры инвариантны при действии транспозиций, а не всей группы S_n . Стоило упомянуть, что для произвольной подстановки, состоящих из нескольких транспозиций, сохранение спектров в общем случае не имеет места.

4. Определение на стр. 45 эквиморфизма групп φ , одного из принципиальных для диссертации понятий, оставляет вопрос: почему невозможно определить его «в одну степень»: $m^a = m^{\varphi(a)}$ вместо $(m^a)^b = (m^{\varphi(a)})^{\varphi(b)}$?

5. Определения 3.2 и 3.3 на стр. 56 лингвистически не согласованы, а появление символа \mathbb{R} усугубляет недоумение.

6. Предлагаемый автором в §3.3 (третьей главы) основной алгоритм анализа джевонс-эквивалентности данных даёт результат для двух фрагментов данных. Несмотря на очевидную ценность полученного решения, едва ли можно говорить о предлагаемом алгоритме как об инварианте группы. Стоило подчеркнуть, что в диссертационном исследовании не ставится задача поиска инварианта группы Джевонса, действующей на множестве булевых функций.

7. Приведенная на стр. 65 формула оценки числа действий сомнений не вызывает, но рекомендаций к оценке значений r_i при этом нет. В главе 4 указанные значения в худшем случае максимальны, либо должны рассчитываться эмпирически. Вероятно, предложенный основной алгоритм имеет полиномиальную временную сложность, и в ряде случаев имеет место равенство $r_i \equiv 1$ (стр. 79). Однако хотелось бы вместо нижнего абзаца стр. 79 иметь если не строгое доказательство этого факта, то точные данные серии численных экспериментов. Возможно, автору в обозримом будущем следует посвятить одну-две публикации изложению и интерпретации своих теоретических результатов и машинных экспериментов в общепринятых терминах эффективности и полиномиальной полноты.

8. В тексте работы присутствуют опечатки и стилистические неточности (выражение «подавляющее меньшинство»; «из общей алгебры верно...» на стр. 20; на стр. 26-27 тремя строками разделены фразы «рассмотрим случай, при котором нет решений» и «поэтому нет смысла рассматривать случаи отсутствия решений», и др.), фразеология естественного языка вытесняется формульными кальками (на стр. 6: «...больше предыдущего в $\dots = 4n$ »; на стр. 79 «...соответствовать значениям $r_i \equiv 1, \forall i$ » и др.), в заключении на стр. 96 внутри абзацев, начинающихся со строчной буквы и заканчивающихся точкой с запятой, помещены фразы, начинающиеся с заглавных букв и заканчивающиеся точками; на стр. 100 в 29-й позиции списка литературы в словосочетании «корреляционно-иммунные» дважды пропущена буква «м» и т.д.

9. Совершенно недостаточен список сокращений и условных обозначений на стр. 97.

Приведённые замечания не сказываются на общей положительной оценке диссертации, а недостатки не влияют на истинность теоретических положений и не снижают значимости полученных автором прикладных результатов. Особо следует отметить тот факт, что в исследовании не только решена теоретически важная и обещающая значительные практические приложения проблема теории информации, но и открыто перспективное направление для дальнейших изысканий.

Заключение по диссертации. Диссертация Кукарцева Анатолия Михайловича соответствует специальности 05.13.17 – Теоретические основы информатики – и является завершённой научно-квалификационной работой, в которой на основании выполненных автором исследований изложены новые научные результаты в области анализа джевонс-эквивалентности данных. Работа выполнена автором самостоятельно на высоком научном и профессиональном уровне. Автореферат корректно отражает содержание диссертации.

Диссертация полностью соответствует требованиям п. 9 «Положения о порядке присуждения учёных степеней» постановления Правительства Российской Федерации № 842 от 24.09.2013 г., а её автор Кукарцев Анатолий Михайлович заслуживает присуждения учёной степени кандидата физико-математических наук.

Д-р техн. наук, доцент, профессор кафедры
Алгебры и математической логики
ФГБОУ ВО НГТУ

Подпись А.В. Чехонадских заверяю:

20 марта 2017 года
г. Новосибирск, ул. Карла Маркса 20,
корпус 8, аудитория 719
Телефон: 8-(383)-346-11-66
e-mail: algebra@nstu.ru



Александр Васильевич