

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ М.В.ЛОМОНОСОВА

На правах рукописи

РЕВЕНКО Артем Викторович

**Построение импликативных зависимостей
для аналитического описания
предметных областей и обнаружения
ошибок в данных**

Специальность 05.13.17 – “Теоретические основы информатики”

АВТОРЕФЕРАТ

**диссертации на соискание ученой степени
кандидата физико-математических наук**

Москва 2013

Работа выполнена на кафедре анализа данных и искусственного интеллекта отделения прикладной математики и информатики ФГБОУ ВПО “Национальный исследовательский университет Высшая школа экономики”.

Научный руководитель доктор физико-математических наук
Кузнецов Сергей Олегович.

Официальные оппоненты: Чечкин Александр Витальевич,
доктор физико-математических наук,
профессор, Военная академия РВСН
имени Петра Великого;
Аншаков Олег Михайлович,
доктор физико-математических наук,
профессор, ФГБОУ ВПО Российский
государственный гуманитарный уни-
верситет.

Ведущая организация ФГБОУ ВПО Воронежский государ-
ственный университет.

Защита состоится “30” октября 2013г. в 16ч. 45м. на заседании диссертацион-
ного совета Д 501.002.16 при ФГБОУ ВПО Московском государственном уни-
верситете имени М.В.Ломоносова по адресу: Российская Федерация, 119991,
Москва, ГСП-1, Ленинские горы, д.1, МГУ имени М.В.Ломоносова, Механико-
математический факультет, аудитория 14-08.

С диссертацией можно ознакомиться в Фундаментальной библиотеке ФГБОУ
ВПО МГУ имени М.В.Ломоносова (Москва, Ломоносовский про-спект, д.27,
сектор А).

Автореферат разослан “30” сентября 2013 г.

Ученый секретарь диссертационного совета
Д 501.002.16 при ФГБОУ ВПО МГУ,
доктор физико-математических наук,
профессор

Корнев Андрей Алексеевич

Общая характеристика работы

Актуальность

В последние десятилетия активно развиваются такие методы и средства автоматизации процессов работы с формальными теориями, как нахождение и проверка правильности доказательств. Однако до нахождения доказательств теорем и до проверки их справедливости необходимо породить гипотезы, которые составляют базу для последующей формулировки теорем. Методы и средства для порождения гипотез в имплицативной форме хорошо известны и изучены в Анализе Формальных Понятий (АФП). Методология, используемая в АФП, может найти применения в самых разных областях науки и техники. Однако существующие методы не являются достаточно гибкими для применения и процесс автоматизации такого исследования недостаточно хорошо изучен и проиллюстрирован, что препятствует дальнейшему распространению АФП. Хорошо известно, что методы АФП являются очень чувствительными к ошибкам. Для расширения области применения таких методов необходимо предложить инструментарий для нахождения ошибок. Решение описанных выше задач позволит, таким образом, разработать инструментарий для автоматического порождения гипотез и подготовить АФП для применения в новых предметных областях.

Цели работы

Целью исследования, результаты которого представлены в диссертации, является разработка методов и средств автоматизации исследования имплицативных зависимостей на основе методологии АФП.

Задачи, решению которых посвящены исследования:

1. разработать методы и средства автоматизации построения имплицативных теорий на примерах:
 - свойств функций на конечных множествах;

- алгебраических тождеств;
2. разработать методы и средства нахождения ошибок в бинарных данных с помощью импликативных зависимостей.

Объектом исследования являются импликативные зависимости.

Предметом исследования являются методы и программные средства автоматизации процессов порождения и анализа импликативных зависимостей.

В соответствии с паспортом специальности 05.13.17 “Теоретические основы информатики” работа включает в себя

- исследования процессов создания, накопления и обработки информации; методов преобразования информации в данные и знания;
- исследования методов преобразования информации в данные и знания;
- создание и исследование информационных моделей, моделей данных и знаний, методов работы со знаниями, методов машинного обучения и обнаружения новых знаний;
- исследования принципов создания и функционирования аппаратных и программных средств автоматизации указанных процессов.

В исследованиях затронуты следующие области из паспорта специальности 05.13.17 “Теоретические основы информатики”.

- 4 Исследование и разработка средств представления знаний; принципы создания языков представления знаний, в том числе – для плохо структурированных предметных областей и слабоструктурированных задач; разработка интегрированных средств представления знаний; средств представления знаний, отражающих динамику процессов, концептуальных и семиотических моделей предметных областей.
- 5 Разработка и исследование моделей и алгоритмов анализа дан-

- ных, обнаружения закономерностей в данных и их извлечения.
- 8 Исследование и когнитивное моделирование интеллекта, включая моделирование поведения, моделирование рассуждений различных типов, моделирование образного мышления.
 - 12 Разработка математических, логических, семиотических и лингвистических моделей и методов взаимодействия информационных процессов, в том числе на базе специализированных вычислительных систем.
 - 14 Разработка теоретических основ создания программных систем для новых информационных технологий.

Основные результаты

В рамках работы автором получены следующие основные результаты.

1. Разработаны и программно реализованы методы автоматизированного изучения импликативных взаимосвязей между свойствами функций на множествах. Испытания программных средств подтвердили их работоспособность.
2. Построено множество импликативных взаимосвязей между свойствами функций на двух-, трех- и четырехэлементных множествах. Получены доказательства части импликаций из базиса импликаций, составлен список недоказанных и не опровергнутых импликаций (открытых вопросов).
3. Разработаны и программно реализованы: методы автоматизированного построения импликативных взаимосвязей между алгебраическими тождествами, алгоритмы проверки выполнения алгебраических тождеств; нахождения алгебр на бесконечном носителе, удовлетворяющих некоторому множеству тождеств и не удовлетворяющих некоторому одному тождеству. Испытания программных средств подтвердили их работоспособность.

4. Построено множество всех истинных импликативных взаимосвязей между алгебраическими тождествами длины не более пяти. Получены доказательства всех импликаций из базиса импликаций.
5. Разработан метод нахождения ошибок в бинарных данных. Получено доказательство того, что разработанный метод удовлетворяет предъявляемым к нему требованиям. Программно реализован интерактивный полиномиальный алгоритм нахождения ошибок в бинарных данных. Проведенные тестовые испытания подтвердили его работоспособность.

Научная новизна

Разработанные автором диссертации методы автоматизации порождения импликативных теорий реализуют новые функциональные возможности. Решена задача нахождения алгебр на бесконечном носителе, заданных некоторым набором тождеств. Автором диссертации предложен новый метод нахождения ошибок в бинарных данных, основанный на АФП. Доказано, что этот метод обеспечивает решение задачи нахождения признаков (или классификации) за полиномиальное время.

Практическая ценность

Методы и средства автоматизации построения импликативных теорий и разработанный метод нахождения ошибок в бинарных данных имеют решающее значение для разработки системы анализа и обработки данных на основе АФП. Разработанные в диссертации методы могут быть использованы для автоматизации построения импликативных теорий в некоторых формализуемых областях наук, таких как, например, алгебраические тождества. Полученные в рамках проведения диссертационного исследования импликативные теории могут быть использованы в соответствующих предметных областях.

Разработанный метод нахождения ошибок в бинарных данных допускает его использование на больших наборах данных.

Апробация

Результаты проведенных исследований докладывались на семинаре кафедры анализа данных и искусственного интеллекта НИУ-ВШЭ (с 2010г. по 2013г., неоднократно), на семинаре института алгебры Политехнического Университета Дрездена (с 2011г. по 2013г., неоднократно) и на семинаре “Проблемы современных информационно-вычислительных систем” на механико-математическом факультете МГУ имени М.В. Ломоносова (2013г.). Результаты докладывались на следующих международных семинарах и конференциях:

1. 19-21.10.2010 Concept Lattices and Their Applications 2010, Sevilla, Spain
[http://www.glc.us.es/cla2010/;](http://www.glc.us.es/cla2010/)
2. 23-25.02.2011 inFormal Concept Analysis Workshop, Dresden, Germany
http://tu-dresden.de/die_tu_dresden/fakultaeten/fakultaet_mathematik_und_naturwissenschaften/fachrichtung_mathematik/institute/algebra/conferences/2011ifca;
3. 11-13.07.2012 European PhD Program in Computational Logic Workshop 2012, Dresden, Germany
[http://www.epcl-study.eu/content/phdws12/;](http://www.epcl-study.eu/content/phdws12/)
4. 1-3.07.2013 European PhD Program in Computational Logic Workshop 2013, Dresden, Germany
<http://www.epcl-study.eu/content/phdws13/>
5. 2.05.2013 IPID Doktorandentreffen, Cologne, Germany
<https://www.daad.de/hochschulen/internationalisierung/ipid/13303.de.html>
6. 28.06.2013 International Seminar, Dresden Germany

http://tu-dresden.de/die_tu_dresden/fakultaeten/fakultaet_mathematik_und_naturwissenschaften/fachrichtung_mathematik/institute/algebra/veranstaltungen;

7. 28.08.2012 What can FCA do for Artificial Intelligence? 2012, Montpellier, France
<http://fca4ai.hse.ru/2012/>;
8. 11-14.10.2012 Concept Lattices and Their Applications 2012, Malaga, Spain
<http://www.matap.uma.es/cla2012/CLA2012/Welcome.html>;
9. 24.03.2013 Formal Concept Analysis meets Information Retrieval 2013, Moscow, Russia
<http://fcair.hse.ru/>;
10. 03.08.2013 What can FCA do for Artificial Intelligence? 2013, Beijing, China
<http://fca4ai.hse.ru/2013>.

По теме диссертации опубликовано 7 печатных работах, в том числе две из них [5, 7] в журналах, включенных в список ВАК.

Структура и объем диссертации

Работа состоит из пяти глав, включая введение, заключение, список литературы. Объем диссертации – 153 страницы, включая оглавление. Список литературы включает 86 наименования. В текст диссертации входят 16 иллюстраций, 2 таблицы, 1 листинг и 3 алгоритма.

Содержание работы

Во **введении** описываются цели работы, обосновывается ее актуальность и практическая значимость. Перечисляются основные результаты, представляются конспективные сведения по истории во-

проса, которые затрагивают положения диссертационного исследования. Дается краткий обзор основных смежных научных направлений. Проводятся сравнения с другими методами анализа данных, использующими алгебраические решетки, а также приводятся примеры применения методов и средств анализа формальных понятий (АФП) для решения задач анализа данных в различных областях наук.

Методология анализа формальных понятий используется в данной работе как основной инструментарий для порождения, анализа и представления имплицативных теорий. Кроме того, разработанные в рамках данной работы методы дополняют методы АФП, особенно в части, которая касается нахождения ошибок в данных.

Во **второй главе** дается введение в АФП, содержание которого предназначено для облегчения понимания положений диссертации. Оно начинается с изложения элементов теории решеток, так как исторически АФП возник при попытке придать прикладную направленность этой теории. В следующем разделе представлено введение в АФП. В двух заключительных подразделах второй главы рассматриваются частные вопросы АФП, имеющие прямое отношение к данной работе, такие как метод активного обучения “Исследование признаков” и вопросы представления знаний в виде диаграммы решетки.

В том числе даются следующие основные определения.

Пусть даны множества G, M . Пусть $I \subseteq G \times M$ - бинарное отношение между множествами G и M . Тройка $\mathbb{K} := (G, M, I)$ называется (*формальным*) *контекстом*. Множество G называют множеством *объектов*, множество M - множеством *признаков*. Пусть $A \subseteq M$, $X \subseteq G$. Рассмотрим отображения $\varphi: 2^G \rightarrow 2^M$ и $\psi: 2^M \rightarrow 2^G$:

$$\begin{aligned}\varphi(X) &:= \{t \in M \mid gIt \text{ для всех } g \in X\}; \\ \psi(A) &:= \{g \in G \mid gIt \text{ для всех } t \in A\}.\end{aligned}$$

Обычно вместо φ и ψ используют единое обозначение $(\cdot)'$. Пусть $Z \subseteq M$ или $Z \subseteq G$. $(Z)''$ называется *замыканием* Z в \mathbb{K} . Пусть

$X \subseteq G, A \subseteq M$. (Формальное) понятие есть пара (X, A) :

$$X' = A \text{ и } A' = X.$$

Пусть $y = (X, A)$. X называется объемом понятия y , A называется содержанием понятия y . Формальные понятия частично-упорядочены отношением

$$(X_1, A_1) \geq (X_2, A_2) \iff X_1 \supseteq X_2 \quad (A_2 \supseteq A_1).$$

Пусть $t \in M, X \subseteq G$, тогда \bar{t} называется отрицательным признаком. $\bar{t} \in X'$ в том и только в том случае, если ни один $x \in X'$ не удовлетворяет xIt . Пусть $A \subseteq M; \bar{A} \subseteq X'$ в том и только в том случае, если все $t \in A$ удовлетворяют $\bar{t} \in X'$. Импликацией в контексте $\mathbb{K} = (G, M, I)$ называется пара (A, B) , пишут $A \rightarrow B$, где $A, B \subseteq M$. A называется посылкой, B называется следствием импликации $A \rightarrow B$. Импликация $A \rightarrow B$ удовлетворяется множеством признаков N , если $A \not\subseteq N$ или $B \subseteq N$. В противном случае считают, что множество признаков противоречит импликации. Импликация $A \rightarrow B$ верна в \mathbb{K} , если она удовлетворяется всеми $g', g \in G$, то есть каждый объект, обладающий всеми признаками из A также обладает всеми признаками из B . Поддержкой импликации в контексте \mathbb{K} называют множество всех объектов контекста \mathbb{K} , чьи содержания содержат как посылку, так и следствие импликации. Атомарной импликацией называют такую импликацию, в следствии которой содержится только один признак, то есть $A \rightarrow b$, где $A \subseteq M, b \in M$. Новые верные импликации из уже имеющихся можно получить с помощью правил Армстронга:

$$\overline{A \rightarrow A}, \quad \frac{A \rightarrow B}{A \cup C \rightarrow B}, \quad \frac{A \rightarrow B, B \cup C \rightarrow D}{A \cup C \rightarrow D}.$$

Базисом импликаций контекста \mathbb{K} называют множество \mathfrak{L} импликаций, из которого любая верная в контексте \mathbb{K} импликация может быть выведена (с помощью правил Армстронга) и никакое собственное подмножество множества \mathfrak{L} не обладает таким свойством. Подмножество признаков $P \subseteq M$ называется псевдосодержанием,

если $P \neq P''$ и для любого псевдосодействия Q , такого что $Q \subset P$, имеет место $Q'' \subset P$. Минимальный по числу импликаций базис известен как *канонический базис импликаций*. Посылки этого базиса можно описать в терминах псевдо-содержаний. Канонический базис импликаций тогда можно представить в виде: $\{P \rightarrow (P'' \setminus P) \mid P - \text{псевдо-содержание}\}$.

Зависимостями предметной области называются такие импликации, которые являются истинными как аксиомы или теоремы в предметной области.

В третьей и четвертой главах на примерах двух предметных областей представлены разработанные автором методы и средства автоматизации, которые необходимы при использовании методологии АФП и в процессе построения импликативных теорий.

Третья глава посвящена решению задачи, направленной на создание методов и средств автоматизации построения взаимосвязей между свойствами функций на конечных множествах. Вопросы о взаимосвязях свойств функций на конечных множествах возникают в разных областях математики. К их числу относятся: теория выбора; теория решеток; теория множеств; примеры исследований взаимосвязей между свойствами функций, которые представлены далее, в разделе 3.1. Однако кроме актуальности для других научных дисциплин, рассмотренная в третьей главе задача обладает еще одним важным преимуществом. В связи с тем, что свойства функций определены математически, представляется целесообразным автоматизировать проверку наличия тех или иных свойств у конкретных функций. Если, кроме того, автоматизировать генерацию функций на конечных множествах, то, тем самым, будет автоматизирован процесс построения начального контекста для исследования признаков.

Как отмечается во второй главе, для применения метода “исследование признаков” необходимо иметь возможность находить контрпримеры к исследуемым импликациям. В данном исследовании контрпримеры были найдены следующим образом. Сначала ге-

нерировался исходный контекст, и строился его канонический базис импликаций. После этого генерировались функции на множестве заданного размера. Эти функции проверялись относительно ранее выбранных свойств, далее производилась проверка полученных комбинаций свойств на удовлетворение каждой импликации. В соответствии с описанием метода в разделе 2.2.1 если хотя бы один контрпример был найден, то этот контрпример добавлялся к контексту, и базис вычислялся заново.

В процессе исследования признаков на свойствах функций не было необходимости проверять все импликации. Благодаря тому, что вопросы взаимосвязей свойств функций имеют длительную историю изучения, некоторые импликации между свойствами функций, получение которых являются целью данной главы, были уже исследованы другими авторами и доказаны как теоремы. Эти знания использовались в ходе проведения данного исследования. Большая часть базисов импликаций на свойствах функций, однако, не была ранее освещена в литературе. Они впервые рассмотрены и частично доказаны в рамках работы в разделе 3.3. Доказательства импликаций, выполненные автором в работе (утверждения 3.3.7-9), а также рядом других авторов (для более полного представления), изложены в разделе 3.3.2. В разделе 3.3.3 приведен подготовленный автором список импликаций, доказать которые не удалось. Отмечено, что в случае доказательства всего представленного списка импликаций, построение импликативной теории будет завершено.

При доказательстве импликаций большую трудность представляли собой импликации, имеющие большую посылку. Практически импликации, имеющие более четырех свойств в посылке, крайне трудны для рассмотрения потому, что многие из этих свойств являются зависимыми. Для упрощения работы с такими импликациями в разделе 3.2 было сформулировано, доказано и в последующем применено утверждение о возможности изменения вида базиса импликаций.

Утверждение 3.2.1 Пусть \mathfrak{B} – канонический базис импликаций контекста \mathbb{K} , содержащий импликацию $H \rightarrow H'' \setminus H$. Тогда, если

$X \subseteq H$ и $X'' = H''$, то множество импликаций \mathfrak{G}_0 , построенное из \mathfrak{G} заменой импликации $H \rightarrow H'' \setminus H$ на импликацию $X \rightarrow H'' \setminus H$, остаётся минимальным по размеру базисом импликаций контекста \mathbb{K} .

В рамках исследования автором доказательно установлена истинность следующих импликаций (утверждения 3.3.3-9):

- ЭКС \rightarrow О, Н, ОАОб;
- ИНТ \rightarrow АОб, Об;
- М \rightarrow В, С;
- Б \rightarrow ИНТ, К, ИД, НП, О, С, Н, М;
- НП \rightarrow ИД, КГ;
- Н \rightarrow ОАОб;
- ИД, В \rightarrow О;
- ИД, М \rightarrow О;
- ИД, КГ \rightarrow НП;
- К \rightarrow О, Н, С;
- ИНТ, НП \leftrightarrow ИНТ, КГ;
- ИНТ, ИД, В \leftrightarrow ИНТ, О;
- ИНТ, НП \leftrightarrow ИНТ, Н, О;
- ЭКС, НП \leftrightarrow ЭКС, ИД, М;
- ИНТ, Н, М \leftrightarrow ИНТ, Б.

Составлен список импликаций, истинность которых предстоит проверить для завершения исследования (открытые вопросы):

1. КГ \rightarrow В;
2. ИД, Об, АОб \rightarrow ОАОб;
3. АТ \rightarrow Н, Об, ОАОб, АОб, В;
4. ИНТ, Об \rightarrow Н, ИД, ОАОб;

5. Н, ИНТ \rightarrow ИД;
6. АТ, С \rightarrow КГ;
7. О, ОБ, АОБ \rightarrow С;
8. Н, ИД, О, АОБ \rightarrow С;
9. О, В, ОБ \rightarrow С;
10. О, АОБ, КГ \rightarrow ОАОБ, С;
11. К \rightarrow ОАОБ, АОБ;
12. НП \rightarrow ОАОБ, С, В, О;
13. Б \rightarrow ОБ;
14. М, ОБ \rightarrow О;
15. ИНТ, КГ \rightarrow Н, К, ОАОБ, С, О;
16. Н, М \rightarrow О;
17. Н, ИД, В \rightarrow С;
18. ИНТ, В, ОБ \rightarrow К, НП, КГ;
19. ИД, КГ \rightarrow ОАОБ, С;
20. М, АОБ, ОБ \rightarrow ОАОБ;
21. ИД, М, ОБ \rightarrow ОАОБ;
22. М, АТ \rightarrow ИД, НП;
23. АТ, ЭКС \rightarrow ИД, НП, М;
24. М, ЭКС \rightarrow КГ;
25. НП, АТ \rightarrow М;
26. НП, ОБ \rightarrow Н;
27. М, К \rightarrow ИД;
28. Н, М, ОБ \rightarrow КГ;
29. М, ЭКС, К \rightarrow Б, ИНТ.

Получена следующая диаграмма решетки рассмотренных

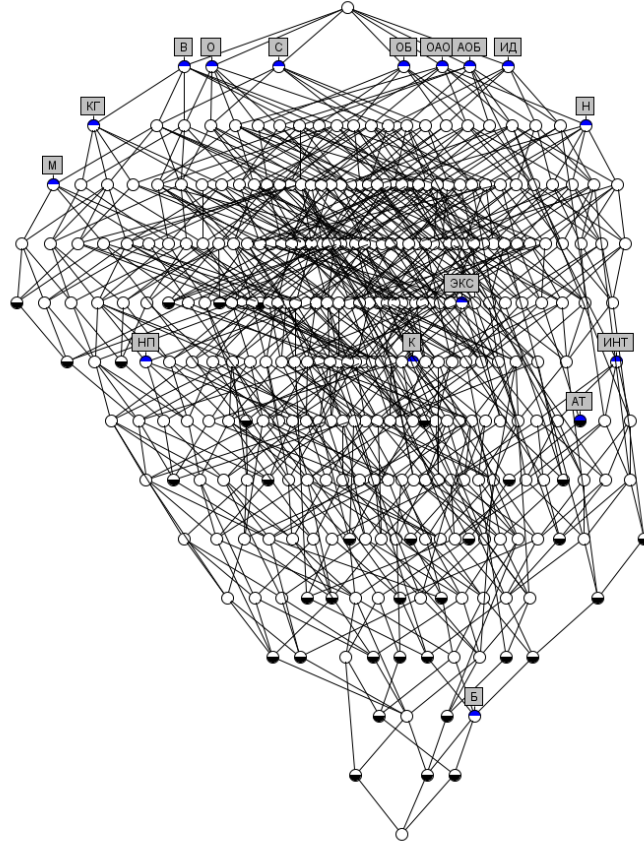


Рис. 1: Диаграмма решетки контекста функций на множествах M_2 , M_3 и M_4

свойств функций на множествах, рис. 1.

В процессе исследования основное внимание было уделено функциям на множествах размера 2 и 3. В разделе 3.3.1 представлены результаты проведенного автором исследования различий в импликативных теориях взаимосвязей свойств функций на множества размера 2 и 3. Сформулированы и доказаны утверждения, объясняющее все различия в импликативных теориях этих типов функций.

Утверждение 3.3.1 *Для любой функции f , определённой на множестве M_2 , из наличия свойства O следует наличие свойства C .*

Утверждение 3.3.2 *Для любой функции f , определённой на множестве M_2 , из наличия свойства ИНТ следует наличие свойства C .*

В разделе 3.4 показано, каким образом можно обобщить определение свойств функций, а вместе с тем – и полученные результаты, на случай функций, заданных не на произвольных множествах, а на решетках.

В результате исследований, представленных в третьей главе, получены следующие результаты.

1. Разработаны методы автоматизации и структуризации процедуры исследования импликативных взаимосвязей свойств функций на множествах. Исследуемые взаимосвязи структурированы в виде решетки понятий, а также в виде базиса импликаций. Представлены доказательства импликаций, подготовлен список открытых импликаций.
2. Разработан и программно реализован метод, позволяющий изменять вид базиса импликаций и отобразить импликации в более удобной для практических приложений форме.
3. Сформулированы и доказаны различия между импликативными теориями для функций на двух- и трехэлементных множествах.
4. Дано обобщение результатов исследования свойств функций на случай функций, заданных на решетках.
5. Разработаны и программно реализованы методы, позволяющие автоматизировать процесс получения импликативных взаимосвязей между свойствами функций. Написанное программное обеспечение удовлетворяет сформулированным требованиям.

В качестве предметной области, результаты исследования которой представлены в **четвертой главе**, были выбраны алгебраические тождества, представляющие, возможно, центральный интерес для универсальной алгебры. В данном исследовании требовалось

добиться еще большей степени автоматизации, чем при изучении свойств функции, так как количество представляющих интерес для приложений тождеств в этом случае значительно больше. Для рассмотрения были выбраны тождества, длина которых не превышает 5 (длиной тождества называется сумма числа вхождений переменных и операций). Таких тождеств насчитывается 70.

Область науки, изучающая общие закономерности алгебраических систем, называется универсальной алгеброй. *Эквациональными классами* в универсальной алгебре называют классы алгебр, удовлетворяющих заданным алгебраическим тождествам. Эквациональные классы играют в этой области центральную роль. В разделе 4.1 приведена ключевая теорема универсальной алгебры, показывающая эквивалентность многообразий алгебр и эквациональных классов.

В разделе 4.2 описаны рассматриваемые в четвертой главе тождества. Алгебраические тождества, как и свойства функций на множествах, заданы строго математически. Более того, проверку удовлетворения конкретной алгеброй конкретного тождества можно автоматизировать. Порождение всех алгебр на носителе ограниченного размера также можно автоматизировать. Эти задачи были решены автором и представлены в разделе 4.3.1. Изоморфизм алгебр позволяет избежать порождения тех алгебр, которые заведомо не могут принести новой информации.

Изучение алгебраических тождеств является подходящей задачей, решение которой целесообразно с применением метода “исследование признаков”, в том числе потому, что на настоящее время относительно высокого уровня развития достигли инструментальные средства автоматического доказательства теорем и нахождения конечных моделей. Для этих целей в процессе диссертационного исследования использовались программы *Mace4* и *Prover9*, решающие поставленные задачи быстрее других средств, которые могут быть использованы для тех же целей, и имеют удобный интерфейс для доступа к ним из внешних программ.

Доказанные в разделе 4.3.2 утверждения показывают, однако, что порождения лишь алгебр на конечном носителе недостаточно для построения импликативной теории выбранных алгебраических тождеств. Задача нахождения бесконечных алгебр, удовлетворяющих заданным тождествам, крайне трудоемка и в общем случае не имеет решения. Автору удалось найти методы порождения алгебр на бесконечном носителе, необходимые для завершения исследования, а также - программно реализовать эти методы (раздел 4.3.3).

Задача порождения алгебр на бесконечном носителе ставится следующим образом. Пусть дана импликация $A \rightarrow b$, $A \subseteq M_{\text{id}}$, $b \in M_{\text{id}}$. Тогда необходимо найти такую алгебру $\mathcal{A} = (\mathbb{N}, *, -, a)$, которая удовлетворяет всем тождествам из A и не удовлетворяет тождеству b .

Для нахождения таких алгебр была зафиксирована форма бинарной, унарной и нулярной операций. Нулярная операция всегда фиксировалась равной нулю $a = 0$. Унарная операция — искалась в следующем виде:

$$-n = \begin{cases} u_0, & \text{если } n = 0; \\ u_1, & \text{если } n = 1; \\ u_2, & \text{если } n = 2; \\ p \times n + q, & \text{если } n \geq 3, \end{cases}$$

где $u_0, u_1, u_2 \in \mathbb{N}_{<5}$, $p \in \mathbb{N}_{<3}$, $q \in \{0, 1, -1, 2, -2, 3, -3\}$.

Бинарная операция $*$ искалась в следующем виде:

$$m * n = \begin{cases} b_{00}, & \text{если } m = 0 \text{ и } n = 0; \\ b_{01}, & \text{если } m = 0 \text{ и } n = 1; \\ b_{10}, & \text{если } m = 1 \text{ и } n = 0; \\ b_{11}, & \text{если } m = 1 \text{ и } n = 1; \\ b_{02}, & \text{если } m = 0 \text{ и } n = 2; \\ b_{03}, & \text{если } m = 0 \text{ и } n = 3; \\ b_{12}, & \text{если } m = 1 \text{ и } n = 2; \\ b_{13}, & \text{если } m = 1 \text{ и } n = 3; \\ b_{20}, & \text{если } m = 2 \text{ и } n = 0; \\ b_{30}, & \text{если } m = 3 \text{ и } n = 0; \\ b_{21}, & \text{если } m = 2 \text{ и } n = 1; \\ b_{31}, & \text{если } m = 3 \text{ и } n = 1; \\ b_{22}, & \text{если } m = 2 \text{ и } n = 2; \\ c_0 \times m + d_0 \times n + e_0, & \text{если } m = 0 \text{ и } n > 3; \\ c_1 \times m + d_1 \times n + e_1, & \text{если } m = 1 \text{ и } n > 3; \\ c_2 \times m + d_2 \times n + e_2, & \text{если } m > 3 \text{ и } n = 0; \\ c_3 \times m + d_3 \times n + e_3, & \text{если } m > 3 \text{ и } n = 0; \\ c_4 \times m + d_4 \times n + e_4, & \text{если } m > 2 \text{ и } m = n; \\ c_5 \times m + d_5 \times n + e_5, & \text{если } m > 1 \text{ и } n = m + 1; \\ c_6 \times m + d_6 \times n + e_6, & \text{если } m > 2 \text{ и } n = m - 1; \\ c_7 \times m + d_7 \times n + e_7, & \text{если } m > 1 \text{ и } n = m + 2; \\ c_8 \times m + d_8 \times n + e_8, & \text{если } m > 3 \text{ и } n = m - 2; \\ c_9 \times m + d_9 \times n + e_9, & \text{если } m > 4 \text{ и } n \neq m, m \pm 1, m \pm 2, \end{cases}$$

где $b_{00}, b_{01}, b_{10}, b_{11} \in \mathbb{N}_{<4}$,

$b_{02}, b_{03}, b_{12}, b_{13}, b_{12}, b_{13}, b_{20}, b_{30}, b_{22} \in \mathbb{N}_{<6}$,

$c_{0-9}, d_{0-9} \in \{0, 1\}$, $e_{0-9} \in \{0, 1, -1, 2, -2, 3, -3\}$.

Все импликации, к которым не были найдены контрпримеры в заданном виде, были доказаны. Таким образом, такой выбор структуры алгебр позволил найти все необходимые контрпримеры.

Для нахождения алгебр в заданном виде автором диссертации предложен алгоритм поиска с возвратом.

Input: $idls_prem \subseteq M_{id}$, $id_conc \in M_{id}$.

Output: Алгебра \mathcal{A} сигнатуры $(*, -, 0)$, удовлетворяющая тождествам $idls_prem$ и не удовлетворяющая тождеству id_conc .

```

1 while True do
2   for id in idls_prem do
3     sat, field = check_identity_partial( $\mathcal{A}$ , id)
4     if sat = False then
5       | backtrack( $\mathcal{A}$ )
6       | break
7     if sat = None then
8       | update( $\mathcal{A}$ , field)
9       | break
10  else
11    sat, field = check_identity_partial( $\mathcal{A}$ , id_conc)
12    if sat = True then
13      | backtrack( $\mathcal{A}$ )
14      | continue
15    if sat = None then
16      | update( $\mathcal{A}$ , field)
17      | continue
18    if sat = False then
19      | return  $\mathcal{A}$ 

```

Алгоритм 1: *find_algebra*

Построение импликативной теории алгебраических тождеств удалось полностью завершить, доказав все импликации, находящи-

еся в каноническом базисе импликаций построенного контекста. Результаты исследования представлены в разделе 4.4. Программный код, тесты к программному коду, получившийся контекст, базис импликаций, а также доказательства импликаций из базиса импликаций вынесены в приложение к диссертационной работе. Требования, предъявляемые к разработанным и реализованным автором методам, заключались в следующем: построить импликативную теорию 70 выбранных тождеств и получить доказательства всех импликаций из базиса импликаций. Требования были выполнены.

В процессе выполнения исследования импликативной теории алгебраических тождеств достигнуты следующие основные результаты.

1. Получено развитие подходов, ранее используемых для автоматизации процессов построения импликативной теории свойств функций на множествах. Разработаны методы автоматизации построения импликативной теории алгебраических тождеств.
2. Разработаны и программно реализованы алгоритмы, которые позволяют проверить, удовлетворяются ли тождества, а также – порождать неизоморфные алгебры на конечных носителях.
3. Разработан метод и реализован алгоритм нахождения алгебры на бесконечном носителе, удовлетворяющей некоторому множеству тождеств и не удовлетворяющей одному выбранному тождеству.
4. С использованием только программных средств построена импликативная теория для всех 70 алгебраических тождеств длины не более 5. Созданное программное обеспечение удовлетворяет предъявляемым к нему требованиям.

В главах 3 и 4 появление ошибок в содержаниях объектов удалось исключить вследствие того, что математически определены признаки контекстов и алгоритмизирована их проверка. Однако показано, что не во всех случаях такое возможно. Например, информация о связях объектов и признаков, найденная в глобальной сети Интернет, не может считаться надежной и отсутствие ошибок га-

рантировать нельзя. В случае проведения исследований методами АФП в политологии, эксперты не всегда могли сойтись во мнении относительно того или иного объекта и признака.

Еще более остро необходимость решения задачи нахождения ошибок в новых объектах контекста встает тогда, когда предполагается одновременный доступ к контексту несколькими пользователями. В таком случае некоторые пользователи могут попытаться намеренно испортить исследование путем добавления новых объектов, содержащих ошибки. На настоящее время в рамках АФП не предложено ни одного подхода к решению этой задачи. Метод нахождения ошибок в новых объектах контекста мог бы расширить область применения АФП на более широкий круг задач, в том числе – на описанные выше, и, тем самым, способствовать появлению новых возможностей для их решения.

В рамках АФП нет никаких внутренних ограничений на содержания объектов, поэтому любой объект может иметь любое содержание, начиная от пустого множества и заканчивая всем множеством признаков. Такие ограничения могут, однако, быть присущи предметной области, а, следовательно, являются внешними по отношению к АФП. Конечно, ничто не мешает использовать внешние ограничения при добавлении новых объектов к контексту. При этом можно проверять содержание всех новых объектов на предмет удовлетворения внешним ограничениям. Однако в таком случае эксперт должен заранее сформулировать весь набор ограничений, чтобы гарантировать отсутствие ошибок. Таким образом, эксперту необходимо задать логические правила, которые описывают все невозможные комбинации признаков в предметной области. Такая задача крайне сложна для решения и, в общем случае, она сравнима с построением всей имплекативной теории вручную. Кроме того, при подобном подходе не учитывается специфика описания предметной области, принятая в АФП, а также уже накопленные знания (объекты, которые уже введены в контекст). По этой причине в рамках данной работы упомянутый подход не рассматривается.

Сущности в АФП задаются прагматическим образом, а именно

– через взаимодействие с другими сущностями. По этой причине любая информация, полученная из названия объекта, будет внешней по отношению к АФП. Любой подход, основанный на использовании названия объекта, не может быть общеприменимым для всех формальных контекстов. Таким образом, не может ставиться задача нахождения верного описания, исходя из названия объекта. Все, что при таком подходе можно гарантировать, это отсутствие такой комбинации формальных признаков в содержании формального объекта, которая не встречается ни в одном реальном объекте из предметной области. Такого условия, однако, достаточно для построения имплицативной теории предметной области, так как сами названия объектов не играют для нее никакой роли.

С другой стороны, как отмечалось в разделе 2.2, множество всех верных импликаций эквивалентным образом представляет формальный контекст. Как следствие, для учета уже накопленных знаний при нахождении ошибок можно использовать множество всех верных импликаций контекста. Кроме того, работа с импликациями для пользователя в общем проще и интуитивно понятнее, так как изначально импликации призваны строго математически формализовать условное высказывание, привычное в обычной жизни. Для пользователя легче ответить на вопрос, выполняется ли импликация в предметной области, чем искать неверные значения в таблице или записывать все возможные правила о том, чего не может быть.

В других областях наук, отличных от АФП, задача нахождения ошибок имеет длинную историю изучения. К сожалению, разработанные методы неприменимы в АФП из-за специфики работы с бинарными таблицами. Краткое описание существующих подходов, а также обоснование необходимости разработки нового метода в рамках АФП, приведено в разделе 5.1.

В **пятой главе** впервые приводится классификация возможных ошибок в содержаниях формальных объектов через зависимости предметной области (раздел 5.2.1). Выделено четыре возможных типа ошибок, являющиеся следствием нарушения зависимостей предметной области.

1. $A \rightarrow b$;
2. $A \rightarrow \bar{b}$;
3. $A \rightarrow b \vee c$;
4. $A \rightarrow \Phi$, где Φ - любая логическая формула кроме представленных выше, например, $\Phi = a \vee (b \wedge \bar{c})$.

Далее рассматриваются только типы 1 и 2, так как они являются наиболее простыми и часто встречающимися.

В разделе 5.2.2 автором на основе проведенной классификации возможных ошибок предложены два разных подхода к нахождению ошибок в содержаниях формальных объектов. Один из них основан на использовании базиса импликаций формального контекста, другой – на вычислении замыканий подмножеств содержания объекта. Так как замыкание может быть вычислено быстро, большее внимание уделено второму подходу. Изначально оба подхода позволяют находить только один тип ошибок. Для подхода, использующего замыкания для выявления ошибок, автору удалось предложить модификацию, позволяющую использовать подход для нахождения еще одного типа ошибок. Модификация, доказательство корректности получаемых результатов, а также алгоритм работы представлены в разделе 5.2.4.

Утверждение 5.2.1 Пусть $\mathbb{K} = (G, M, I)$, $A \subseteq M$. Множество

$$\mathcal{I}_A = \{B \rightarrow d \mid B \in \mathcal{MC}_A, d \in B'' \setminus A \cup \overline{A \setminus B}\},$$

где $\mathcal{MC}_A = \{B \in \mathcal{C}_A \mid \nexists C \in \mathcal{C}_A : B \subset C\}$ и $\mathcal{C}_A = \{A \cap g' \mid g \in G\}$, является таким множеством всех атомарных импликаций (или их следствий с некоторыми признаками, добавленными в посылку), имеющих вид зависимостей Типов 1 и 2, что они верны в контексте \mathbb{K} , не удовлетворяются A и имеют ненулевую поддержку.

Алгоритм выглядит следующим образом.

Input: $\mathbb{K} = (G, M, I)$, $A \subseteq M$.
Output: Импликации, которые противоречат A .

```

1 if  $A'' = A$  then
2   return  $\emptyset$ 
3 Candidates = {object'  $\cap$  A | object  $\in$  G}
4 Candidates = {C  $\in$  Candidates |  $\nexists B \in$  Candidates: C  $\subseteq$  B}
5 Result  $\leftarrow$   $\emptyset$ 
6 for Candidate in Candidates do
7   Result.add({Candidate  $\rightarrow$  d |
8     d  $\in$  (Candidate''  $\setminus$  A  $\cup$   $\overline{A \setminus$  Candidate)})}
9 return Result

```

Алгоритм 2: inspect_direct

Утверждение 5.2.2 *Временная сложность алгоритма 2 составляет $O(|G|^2 \times |M|^2)$.*

В разделе 5.2.5 сравниваются результаты работы обоих предложенных выше подходов на примере из раздела 5.2.3, на случайных контекстах и на реальных данных. Метод нахождения ошибок, предложенный в настоящей главе, допускает другое более наглядное представление через диаграмму решетки (раздел 5.3). Завершает изложение методов нахождения ошибок в содержаниях формальных объектов раздел 5.4, в котором приведен исходный код с комментариями о его структуре, а также тесты, использованные для проверки работы исходного кода.

В пятой главе автором также предложено одно из возможных нетривиальных применений разработанных методов для отладки исходного программного кода. Подход излагается в разделе 5.5, приводится пример применения метода. В разделе 5.6 представлен исходный код, который использовался для реализации предложенного способа отладки исходного программного кода, а также тесты к нему.

Основными результатами, представленными в пятой главы, являются решения следующих задач.

1. Приведена классификация возможных ошибок в бинарных

данных с помощью представления этих ошибок через зависимости в предметной области.

2. На основе проведенной классификации ошибок предложены два метода нахождения ошибок. Отмечено, что один из них находит результаты за меньшее время. Доказано, что этот метод приводит к нахождению нарушенных взаимосвязей, которые до этого выполнялись хотя бы для одного объекта из контекста.
3. Разработаны модификации более быстрого метода, позволяющие находить ошибки двух классов. Доказана корректность результатов.
4. Разработан и программно реализован интерактивный полиномиальный алгоритм нахождения ошибок в бинарных данных. Получена характеристика в виде модифицированного метода ближайших соседей.
5. Проведены эксперименты по оценке качества нахождения ошибок, а также – по скорости работы на случайных данных и на реальных данных.
6. Основываясь на методе нахождения ошибок предложен способ нахождения ошибок в исходном программном коде.

В рамках исследования импликативной теории свойств функций на множествах разработаны общие подходы к автоматизации процесса такого исследования. В результате дальнейшего развития и улучшения ранее разработанных подходов в рамках исследования алгебраических тождеств автору удалось в автоматическом режиме построить импликативную теорию всех 70 алгебраических тождеств длины не более 5. Необходимо отметить, что результат получен за счет комбинирования прагматического подхода, характерного для АФП, и учета структуры предметной области, позволяющей находить контрпримеры к получаемым импликациям. Разработанные методы нахождения ошибок в бинарных данных позволяют расширить область применения разработанных методов автоматического построения импликативных теорий на тот случай, когда достовер-

ность данных нельзя гарантировать, однако с данными работает эксперт в предметной области.

В заключении представлены основные результаты работы.

1. Разработаны и программно реализованы методы автоматизированного изучения импликативных взаимосвязей между свойствами функций на множествах. Испытания программных средств подтвердили их работоспособность.
2. Построено множество импликативных взаимосвязей между свойствами функций на двух-, трех- и четырехэлементных множествах. Получены доказательства части импликаций из базиса импликаций, составлен список недоказанных и не опровергнутых импликаций (открытых вопросов).
3. Разработаны и программно реализованы: методы автоматизированного построения импликативных взаимосвязей между алгебраическими тождествами, алгоритмы проверки выполнения алгебраических тождеств; нахождения алгебр на бесконечном носителе, удовлетворяющих некоторому множеству тождеств и не удовлетворяющих некоторому одному тождеству. Испытания программных средств подтвердили их работоспособность.
4. Построено множество всех истинных импликативных взаимосвязей между алгебраическими тождествами длины не более пяти. Получены доказательства всех импликаций из базиса импликаций.
5. Разработан метод нахождения ошибок в бинарных данных. Получено доказательство того, что разработанный метод удовлетворяет предъявляемым к нему требованиям. Программно реализован интерактивный полиномиальный алгоритм нахождения ошибок в бинарных данных. Проведенные тестовые испытания подтвердили его работоспособность.

Благодарности

Автор выражает глубокую благодарность научному руководителю, доктору физико-математических наук Сергею Олеговичу Кузнецову за постановку задач, постоянное внимание и помощь в работе. Автор выражает глубокую благодарность профессору Васенину Валерию Александровичу за плодотворные обсуждения, критические замечания и помощь в работе. Автор выражает глубокую благодарность профессору Бернарду Гантеру за плодотворное обсуждение и внимание к работе.

Список публикаций автора

- [1] A. Revenko. Debugging programs using formal concept analysis. In *FCAIR 2013*, pages 105–112. CEUR, 2013, 0.8 п. л. (личный вклад 0.8 п. л.).
- [2] A. Revenko. Debugging program code using implicative. In *FCA4AI 2013*, pages 27–39. CEUR, 2013, 1 п.л. (личный вклад 1 п. л.).
- [3] A. Revenko and S.O. Kuznetsov. Attribute exploration of properties of functions on ordered sets. In *Proc. 7th International Conference on Concept Lattices and Their Applications, University of Sevilla*, pages 311–324, 2010, 0.9 п. л. (личный вклад 0.6 п. л.).
- [4] A. Revenko and S.O. Kuznetsov. Finding errors in new object intents. In *CLA 2012*, pages 151–162. CEUR, 2012, 1 п.л. (личный вклад 0.8 п. л.).
- [5] A. Revenko and S.O. Kuznetsov. Attribute exploration of properties of functions on sets. *Fundamenta Informaticae*, 115(4):377–394, 2012, 1.2 п. л. (личный вклад 0.7 п. л.).
- [6] A. Revenko, S.O. Kuznetsov, and B. Ganter. Finding errors in new object in formal contexts. In *FCA4AI 2012*, pages 65–72. CEUR, 2012, 0.7 п. л. (личный вклад 0.3 п. л.).
- [7] А. Ревенко. Нахождение ошибок в бинарных таблицах данных.

Научно-техническая информация. Серия 2. «Информационные процессы и системы», 6:10–18, 2013, 1 п. л. (личный вклад 1 п. л.).