

ФГБОУ ВПО Московский государственный университет
имени М.В. Ломоносова

На правах рукописи

БОЛОТОВА Светлана Юрьевна

**Разработка и исследование метода релевантного обратного
вывода**

Специальность 05.13.17 – теоретические основы информатики

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата физико-математических наук

Москва – 2014

Работа выполнена на кафедре математического обеспечения ЭВМ факультета прикладной математики, информатики и механики ФГБОУ ВПО «Воронежский государственный университет».

Научный руководитель: доктор физико-математических наук,
доцент
Махортов Сергей Дмитриевич

Официальные оппоненты: доктор физико-математических наук,
ведущий научный сотрудник
Бурдонов Игорь Борисович,
ФГБУН «Институт системного программирования РАН»

доктор технических наук, профессор
Громов Юрий Юрьевич,
ФГБОУ ВПО «Тамбовский государственный
технический университет»

Ведущая организация: ФГБОУ ВПО «Московский государственный
технический университет имени Н.Э. Баумана»

Защита диссертации состоится 19 марта 2014 года в 16 часов 45 минут на заседании диссертационного совета Д 501.002.16, созданного на базе ФГБОУ ВПО «Московский государственный университет имени М.В. Ломоносова», по адресу: 119991, Москва, ГСП-1, Ленинские горы, д.1, ФГБОУ ВПО МГУ имени М.В. Ломоносова, механико-математический факультет, аудитория 14-08.

С диссертацией можно ознакомиться в Фундаментальной библиотеке ФГБОУ ВПО «Московский государственный университет имени М.В. Ломоносова» (Москва, Ломоносовский проспект, д.27, сектор А, 8 этаж).

Автореферат разослан « ____ » _____ 2014 г.

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

Корнев Андрей Алексеевич

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Тематика и актуальность исследования. Информационные технологии на современном этапе развития общества глубоко проникают практически во все сферы деятельности человека, включая материальную, интеллектуальную, социально-культурную и политическую. Подобные тенденции наблюдаются в технике, экономике, медицине, коммерции, образовании, связи, в военной области и так далее. Указанное обстоятельство порождает неуклонно возрастающую зависимость общества от уровня эффективности и надежности информационных систем, обеспечивающих процессы его жизнедеятельности. Информационные системы становятся глобальными, эволюционирующими, социо-критическими, экологически опасными¹. Соответственно возрастают риски чрезвычайных ситуаций, экологических катастроф, материальных и других потерь национальных масштабов².

Ответом на подобные вызовы современности служат применяемые в процессах разработки инновационные решения, повышающие эффективность, надежность и безопасность. К таким решениям могут быть отнесены методы интеллектуализации информационных систем.

Искусственный интеллект, зародившийся как научная дисциплина еще в 50-х годах прошлого столетия, имеет богатую и противоречивую историю. Так и не заменив интеллект человека, он тем не менее стал неотъемлемой частью современных информационных технологий. Интеллектуальная система – это техническая или программная система, способная решать задачи, традиционно считающиеся творческими, принадлежащие конкретной предметной области, знания о которой хранятся в памяти такой системы. Подобные системы характеризуются, в частности, способностью «логически мыслить», то есть осуществлять логический вывод (прямой или обратный) некоторых фактов или решений на основе формальной логики.

Однако формальные логики и соответствующие им процессы логического вывода нередко обладают весьма ресурсоемкой архитектурой, как правило, имеющей экспоненциальный или NP-полный характер, как с точки зрения требуемого объема памяти, так и сложности вычислений. Именно поэтому, несмотря на бурное развитие компьютерной техники, до сих пор более или менее широкое практическое применение получили лишь такие интеллектуальные системы, которые основаны на упрощенных логиках. К таковым, в частности, относятся продукционные экспертные системы.

Интеллектуальные системы продукционного типа представляют важное направление исследований в области искусственного интеллекта. Они используются в теории обучающих систем, систем принятия решений, а также при разработке практических экспертных систем, охватывающих различные

¹ Чечкин А.В. Ультрамножественная модель базы данных и ультраоператорная модель базы знаний проблемной области сложной системы на основе среды нейрорадикалов / А.В. Чечкин // Нейрокомпьютеры: разработка, применение. – 2009. – №12. – С. 4–9.

² Критически важные объекты и кибертерроризм. Часть 1. Системный подход к организации противодействия / О. О. Андреев и др. ; под ред. В. А. Васенина. – М. : МЦНМО, 2008. – 398 с.

прикладные области, такие как медицина, проектирование, геологоразведка и другие.

Роль аппарата продукционно-логических систем как **актуального предмета исследований** повышается и другими факторами. В своей монографии³ Д.А. Поспелов рассматривает системы продукций в широком смысле, как основу моделирования рассуждений в архитектуре ЭВМ, роботах, экспертных системах и т.д. В последние годы был показан и изучен продукционный характер многих различных моделей в информатике, в том числе – непосредственно не относящихся к области искусственного интеллекта. Кроме того, имея относительно несложную структуру, системы продукционного типа могут привлекаться в качестве стартового «полигона» для создания и изучения методов управления формальными знаниями, которые в дальнейшем могут быть применены и для построения более сложных интеллектуальных систем.

С другой стороны, в связи с кардинальным усложнением решаемых задач, повышается ответственность разработчиков информационных систем, увеличивается цена сделанной ошибки. Как следствие, оказывается весьма востребованным создание строгой математической базы, которая теоретически обосновывала бы корректность и надежность таких систем, а также возможности их автоматической оптимизации.

Эффективным средством формального построения и исследования компьютерных программ, основанных на различных парадигмах, являются алгебраические системы. Это положение в полной мере относится к логическому программированию, особенно в части представления знаний.

Математическую основу для создания и исследования моделей знаний предоставляет алгебраическая логика. Однако, расширяя возможности исследования самих логических теорий, она не повышает эффективности их практического применения. В силу универсальности она не решает важных частных проблем, связанных с упомянутыми выше системами продукционного типа. К задачам такого рода относятся эквивалентные преобразования, верификация продукционных и подобных им систем.

В работах С.Д. Махортова⁴ сформулирована основанная на решетках новая алгебраическая теория (LP-структур) как общая модель продукционно-логических систем широкого спектра. Она позволяет с единых позиций рассматривать и успешно решать перечисленные выше задачи. Одно из направлений ее применения – оптимизация обратного логического вывода и связанная с ним верификация знаний. В частности, высказана принципиально новая идея о минимизации числа обращений к внешним устройствам в процессе обратного вывода и указан способ ее реализации⁵ (релевантный LP-вывод).

³ *Поспелов Д.А.* Моделирование рассуждений. Опыт анализа мыслительных актов / Д.А. Поспелов. – М.: Радио и связь, 1989. – 184 с.

⁴ *Махортов С. Д.* Логические отношения на решетках / С. Д. Махортов // Вестник ВГУ. Серия Физика, математика. – Воронеж. – 2003. – № 2. – С. 203–209.

⁵ *Махортов С. Д.* Релевантный обратный вывод и верификация логических программ на основе решения уравнений в LP-структурах / С. Д. Махортов // Методы и средства обработки информации: Третья Всероссийская научная конференция. Москва, 6-8 октября 2009 г.: Труды конференции / Под ред. Л.Н. Королева. – М. : ВМиК МГУ, 2009. – С. 143–148.

Данная идея не пересекается с известными методами быстрого вывода в продукционных системах, а дополняет их. Алгоритмы RETE (Forgy C. L.) и TREAT (Miranker D.) разработаны для стратегии прямого вывода и использовались в продукционных системах с прямым выводом (например, OPS5, CLIPS). Алгоритм LEAPS (Miranker D.) компилирует множество правил продукционной системы OPS5 в язык С. В последующих модификациях наиболее популярный алгоритм RETE адаптировался для смешанного (двунаправленного) вывода (Homeier P. V., Lee Y. H.). Предложенная в статье С.Д. Махортова⁶ и развиваемая в настоящей диссертации концепция использования уравнений в LP-структурах предназначена для оптимизации исключительно *обратного вывода* с точки зрения минимизации обращений к внешним устройствам. Кроме того, можно адаптировать имеющиеся ускоренные алгоритмы обратного вывода для решения рассматриваемых в работе продукционно-логических уравнений.

Далее отметим, что в работах С.Д. Махортова релевантный LP-вывод остался практически на уровне общей идеи. В частности, не были рассмотрены и сопоставлены различные стратегии выбора параметров релевантности. Не было проведено ни достаточного количества тестов, ни, тем более, каких-либо статистических исследований, которые могли бы действительно обосновать и оценить эффективность LP-вывода. Наконец, в представленной ранее его реализации не использовались параллельные вычисления, которые, безусловно, могли еще более повысить быстродействие нового метода.

Таким образом, имеется **актуальная практическая задача** сокращения обмена данными с внешними устройствами и, тем более, интерактивным пользователем. Для ее решения необходимо рассмотреть **научную задачу** разработки и исследования метода релевантного LP-вывода, обеспечивающего ускорение продукционно-логического вывода и верификацию знаний.

Основная цель диссертации – повышение эффективности обратного продукционно-логического вывода и автоматизированных процессов верификации знаний с помощью метода релевантного LP-вывода, основанного на усовершенствованной схеме решения логического уравнения, предложенных способах выбора параметров релевантности и разработанных параллельных алгоритмах.

Перечисленные выше вопросы относятся к области исследования моделей знаний, методов работы со знаниями, а также определения принципов построения программных средств автоматизации управления знаниями. Данные положения непосредственно отражены в формуле и паспорте научной специальности 05.13.17 – Теоретические основы информатики (пп. 4, 8, 14).

Использованные в работе **методы исследований** основаны на теориях множеств, решеток, бинарных отношений, универсальных алгебр, алгебраической логики, теории графов, математической статистики,

⁶ Махортов С. Д. Логические уравнения на решетках / С. Д. Махортов // Вестник ВГУ. Серия Физика, математика. – Воронеж. – 2004, № 2. – С. 170–178.

параллельных вычислений. При описании приложений применяются методы анализа алгоритмов, теории и технологий программирования.

Объектом исследования являются продукционно-логические системы. **Предмет исследования** – процессы обратного логического вывода в продукционных системах.

Научная новизна диссертации заключается в следующих положениях.

- Сформулирована и исследована обобщенная модель LP-структуры, усовершенствована схема процесса решения продукционно-логического уравнения.

- Разработан новый метод обратного вывода в логических системах продукционного типа, ориентированный на снижение числа обращений к внешним устройствам.

- Предложены и апробированы различные способы выбора параметров релевантности для процессов релевантного и кластерно-релевантного LP-вывода.

- Разработаны параллельные алгоритмы релевантного и кластерно-релевантного LP-вывода, а также параллельные алгоритмы вычисления истинных прообразов в LP-структурах.

- Проведено подробное исследование полученных результатов методами математической статистики. Таким образом формально обоснованы преимущества нового метода обратного вывода в различных его модификациях.

Теоретическая и практическая значимость.

Работа имеет в основном теоретический характер. Построенные в ней модели, методы и алгоритмы представляют научно обоснованные технологические решения, позволяющие существенно повысить эффективность обратного продукционно-логического вывода и автоматизированных процессов верификации знаний.

Практическая значимость работы заключается в программной реализации разработанных параллельных алгоритмов, реализующих LP-вывод, и построенной на их основе новой версии программной системы LPExpert, которая обладает эффективными возможностями создания, эксплуатации и верификации продукционно-логических баз знаний для различных предметных областей.

Достоверность представленных результатов обеспечивается строгими математическими формулировками и доказательствами, а также многочисленными тестами и статистическими исследованиями их результатов.

Результаты диссертации докладывались на следующих научных конференциях и семинарах:

- IV Международной научной конференции «Современные проблемы прикладной математики, теории управления и математического моделирования» (ПМТУММ-2011) (Воронеж, 2011);
- III Всероссийской конференции с международным участием «Знания – Онтологии – Теории» (ЗОНТ-2011) (Новосибирск, 2011);

- X Всероссийской научной конференции «Нейрокомпьютеры и их применение» (НКП-2012) (Москва, 2012);
- Международной молодежной конференции-школе «Современные проблемы прикладной математики и информатики» (МРАМС'2012) (Дубна, 2012);
- V Международной научной конференции «Современные проблемы прикладной математики, теории управления и математического моделирования» (ПМТУММ-2012) (Воронеж, 2012);
- Международной конференции «Актуальные проблемы прикладной математики, информатики и механики» (Воронеж, 2012);
- XI Всероссийской научной конференции «Нейрокомпьютеры и их применение» (НКП-2013) (Москва, 2013);
- International Conference “Distributed Intelligent Systems and Technologies (DIST’2013)” (St. Petersburg, July 1–4, 2013);
- International Conference “Mathematical Modeling and Computational Physics” (ММСР’2013) (Dubna, July 8–12, 2013);
- Всероссийской конференции с международным участием «Знания – Онтологии – Теории» (ЗОНТ-2013) (Новосибирск, 2013);
- научном семинаре «Проблемы современных вычислительных систем» механико-математического факультета МГУ имени М.В. Ломоносова, рук. В.А. Васенин (Москва, 2013);

а также научных сессиях Воронежского госуниверситета (2011–2013).

Публикации. По теме диссертации опубликовано 14 научных работ, список которых приведен в ее заключительной части. Статьи [1–4] соответствуют Перечню ВАК РФ. Опубликованные работы вполне отражают содержание диссертации. Получено свидетельство о регистрации компьютерной программы [15].

Личный вклад автора. В диссертационной работе изложены результаты, полученные лично автором, включая исследование проблематики, решения задач, алгоритмы и программную реализацию. Из 4-х совместных работ с научным руководителем в диссертацию вошли только результаты автора.

Структура и объем работы. Диссертация состоит из введения, четырех глав, заключения, двух приложений и списка литературы. Общий объем диссертации 162с., список литературы содержит 125 наименований.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

В автореферате приводится лишь часть сформулированных в работе утверждений, поэтому непрерывность нумерации формул, определений и теорем может не соблюдаться.

Во **введении** содержатся общий обзор тематики исследований и ее ретроспектива, показана актуальность работы, представлены цели и задачи исследований, описаны методы решения задач, сформулированы полученные в

работе новые научные результаты, показаны научное и практическое значение работы.

В **первой главе** излагаются базовые положения теории LP-структур. Первоначально они были введены и обоснованы в работах С.Д. Махортова. В настоящей главе они получили уточнение и развитие. Представленные здесь результаты основаны на более слабых по сравнению с работой⁷ условиях на LP-структуру. Основное изменение связано с использованием вместо решеточного отношения частичного порядка \supseteq , отношения вида \supseteq_R , содержащего лишь необходимое для исследования отношения R подмножество \supseteq в основных определениях, формулировках теорем и доказательствах.

Под *LP-структурой* подразумевается алгебраическая система, представляющая решетку, на которой задано бинарное отношение, обладающее некоторыми продукционно-логическими свойствами.

Отношение называется *продукционно-логическим*, если оно обладает рефлексивностью, транзитивностью и другими свойствами, которые определяются конкретной моделью. Одно из таких свойств – дистрибутивность. Неформально оно означает возможность логического вывода по частям и объединения его результатов на основе решеточных операций.

В настоящей работе в качестве основы LP-структур рассматриваются решетки с семантикой подмножеств – $\lambda(F)$ (множество конечных подмножеств F). Соответственно используются знаки теоретико-множественных операций \subseteq , \supseteq , \cap и \cup . В такой нотации \cup -дистрибутивность трактуется в следующем смысле: из $(A, B_1), (A, B_2) \in R$ следует $(A, B_1 \cup B_2) \in R$. Это свойство называется *дистрибутивностью отношения R* .

Пусть на решетке \mathbb{F} задано некоторое бинарное отношение R . Совокупность всех атомов решетки, содержащихся в элементах пар отношения R , будем называть множеством атомов, которыми оперирует отношение R . Построенное на объединениях этих атомов подмножество исходной решетки обозначим \mathbb{F}_R . Очевидно, \mathbb{F}_R – подрешетка в \mathbb{F} . Для заданного R введем бинарное отношение \supseteq_R – такое подмножество решеточного частичного порядка \supseteq , что элементы всех пар \supseteq_R принадлежат \mathbb{F}_R . Обозначим также \supset_R подмножество \supseteq_R , не содержащее рефлексивных пар (вида (A, A)).

Итак, в рамках настоящей работы действует следующее определение.

Определение 1.2.1. Бинарное отношение на решетке называется продукционно-логическим (или просто *логическим*), если оно содержит \supseteq_R , дистрибутивно и транзитивно.

В отличие от работ С.Д. Махортова, использующих в аналогичной ситуации отношение \supseteq , определение 1.2.1 формулирует *более слабое условие* на определяемое понятие и, следовательно, открывает возможности для получения новых результатов.

⁷Махортов С. Д. Логические отношения на решетках / С. Д. Махортов // Вестник ВГУ. Серия Физика, математика. – Воронеж. – 2003. – № 2. – С. 203–209.

Логическим замыканием заданного на решетке произвольного бинарного отношения R называется наименьшее продукционно-логическое отношение, содержащее R .

Теорема 1.3.1. Для произвольного бинарного отношения R на решетке существует логическое замыкание \xrightarrow{R} .

Теорема о существовании логического замыкания позволяет ввести понятие эквивалентных отношений, что, в свою очередь, может явиться основой для реализации формальных преобразований и оптимизации баз знаний продукционного типа. Два отношения R_1, R_2 называются *эквивалентными*, если их логические замыкания совпадают.

Теорема 1.3.2. Пусть R_1, R_2, R_3, R_4 – отношения на общей решетке. Если при этом $R_1 \sim R_2$ и $R_3 \sim R_4$, то $R_1 \cup R_3 \sim R_2 \cup R_4$.

Она обосновывает принцип локальности эквивалентных преобразований логических отношений: если часть данного отношения заменить эквивалентной частью, то и в целом новое отношение будет эквивалентным исходному. Этот результат открывает возможности автоматических преобразований баз знаний.

Отношение R на атомно-порожденной решетке \mathbb{F} называется *каноническим*, если оно задано множеством пар вида (A, a) , где $A \in \mathbb{F}$, a – атом в \mathbb{F} . Доказано

Утверждение. Для любого отношения на атомно-порожденной решетке существует эквивалентное ему каноническое отношение.

Рассмотрены вопросы, связанные с эквивалентной минимизацией LP-структур. *Логической редукцией* отношения на решетке называется эквивалентное ему минимальное отношение.

Для отношения R на решетке \mathbb{F} рассмотрим отношение \tilde{R} , построенное по R последовательным выполнением следующих шагов:

- добавить к R все пары вида (A, A) , где $A \in \mathbb{F}_R$, и обозначить новое отношение R_1 ;
- добавить к R_1 всевозможные пары (A, B) с элементами вида $A = \bigcup_i A_i, B = \bigcup_i B_i$, где все (A_i, B_i) ($i = 1, \dots, n$) принадлежат R_1 ;
- объединить полученное отношение с отношением включения \supset_R .

Далее для отношения R рассмотрим отношение \underline{R} , построенное по данному R последовательным выполнением шагов, обратных построению \tilde{R} , а именно:

- исключить из R содержащиеся в нем пары вида $A \supset_R B$ и обозначить новое отношение R_{-1} ;
- исключить из R_{-1} всевозможные пары (A, B) с элементами вида $A = \bigcup_i A_i, B = \bigcup_i B_i$, где все (A_i, B_i) ($i = 1, \dots, n$) принадлежат R_{-1} и не совпадают с (A, B) ;
- исключить из полученного отношения все рефлексивные пары.

Следующая теорема указывает достаточное условие существования и способ построения логической редукции данного отношения.

Теорема 1.5.2. Пусть для бинарного отношения R , заданного на решетке \mathbb{F} , построено соответствующее отношение \tilde{R} . Тогда, если для \tilde{R} существует транзитивная редукция R^0 , то соответствующее ей отношение \tilde{R}^0 представляет собой логическую редукцию исходного отношения R .

Изложенные выше результаты содержат существенное усиление по сравнению с аналогичными теоремами работы⁸, основанными на использовании в определении логических отношений операции \supseteq . В частности, требование существования транзитивной редукции отношения \tilde{R} при определенных условиях может оказаться избыточным для существования логической редукции исходного отношения R . Очевидно, что при конечном множестве R из него всегда можно последовательно удалить «лишние» пары, получив в результате логическую редукцию. Однако, если при этом сама решетка \mathbb{F} бесконечна и имеет соответствующую структуру, то, объединяя R с отношением \supseteq , можно получить не имеющее транзитивной редукции отношение \tilde{R} . Таким образом, использование отношения \supseteq_R способно исправить ситуацию.

В заключительном разделе главы 1 вводится связанный с LP-структурами класс логических уравнений, порожденный логическим замыканием канонического отношения R на решетке.

Определение 1.6.1. Атом x решетки \mathbb{F} называется начальным при отношении R , если в R нет ни одной такой пары (A, B) , что x содержится в B и не содержится в A . Элемент X решетки \mathbb{F} называется начальным, если все его атомы являются начальными при отношении R . Подмножество $\mathbb{F}_0(R)$ решетки \mathbb{F} , состоящее из всех начальных элементов \mathbb{F} , называется начальным множеством решетки \mathbb{F} (при отношении R).

Обозначим R^L – логическое замыкание отношения R и рассмотрим уравнение

$$R^L(X) = B, \tag{1}$$

где $B \in \mathbb{F}$ – заданный элемент, $X \in \mathbb{F}$ – неизвестный.

Определение 1.6.2. Частным решением X уравнения (1) называется любой минимальный прообраз элемента B , содержащийся в \mathbb{F}_0 . Приближенным (частным) решением X уравнения (1) называется любой прообраз элемента B , содержащийся в \mathbb{F}_0 . Общим решением уравнения (1) называется совокупность всех его частных решений $\{X_s\}$, $s \in S$.

Приводятся результаты о разрешимости и количестве решений этих уравнений, а также излагаются методы их решения. Нахождение решения логического уравнения соответствует обратному логическому выводу на решетке. Изложенные в данном разделе результаты были ранее получены в

⁸ Махортов С. Д. Логические отношения на решетках / С. Д. Махортов // Вестник ВГУ. Серия Физика, математика. – Воронеж. – 2003. – № 2. – С. 203–209.

работе⁹ для более частной LP-структуры. Их доказательства в нашем случае существенно не изменились, поэтому в работе не приводятся. Однако внесено изменение в пошаговую схему решения уравнения. В настоящей работе разбиение исходного отношения на слои производится сразу, до упрощения правой части уравнения. В результате из процесса решения исключается избыточный этап, связанный с идентификацией приближенных решений.

Терминология, концепция уравнений и механизмы их решения лежат в основе методов релевантного LP-вывода, составляющих основной предмет настоящего диссертационного исследования.

Вторая глава посвящена методологии релевантного и кластерно-релевантного обратного вывода. Она направлена на снижение числа обращений к внешним источникам информации. Предложены, обоснованы и апробированы различные способы выбора параметров релевантности. Дано описание алгоритмов решения логических уравнения в слоях, а также алгоритмов исследования начальных прообразов на предмет истинности. Далее описаны алгоритмы параллельного решения указанных задач. Результаты могут быть также применены для верификации соответствующих баз знаний.

Понятие LP-структуры представляет абстрактное описание теоретико-множественной модели различных классов продукционно-логических систем. В качестве конкретной предметной области в работе рассматривается известный тип продукционных экспертных систем¹⁰. Он выбран в силу своей простоты и широкой известности, что позволяет непосредственно сосредоточиться на изложении основных идей и методов релевантного вывода. Однако рассматриваемые подходы могут быть применены к системам с более сложной логикой, в том числе неклассическим.

В рассматриваемой модели объектами бинарного отношения R являются конечные подмножества исходного множества F атомарных фактов. Состоящая из таких объектов математическая структура представляет собой частный случай решетки $\mathbb{F} = \lambda(F)$.

Аппарат логических уравнений предоставляет возможности для ускорения обратного вывода в продукционной системе. Следует отметить, что в случае продукционных систем оценки вычислительной сложности алгоритмов логического вывода неудовлетворительны. Действительно, имея множество N возможных атомарных фактов, можно сформировать до 2^N предпосылок и заключений потенциальных правил, что соответствующим образом скажется на оценках сложности алгоритмов машины вывода. Таким образом, важное значение приобретают альтернативные способы повышения эффективности.

В частности, существенную роль играет место хранения фактов. В то время как множество правил обычно может целиком располагаться в основной памяти компьютера, для начальных фактов такое хранение возможно далеко не всегда. В то же время единственная операция чтения с диска или, более того, получения

⁹ Махортов С. Д. Логические уравнения на решетках / С. Д. Махортов // Вестник ВГУ. Серия Физика, математика. – Воронеж. – 2004, № 2. – С. 170–178.

¹⁰ Sowyer B. Programming Expert Systems in Pascal / B. Sowyer, D. Foster. –John Wiley & Sons, Inc., 1986.

информации от интерактивного пользователя, по времени способна перекрыть сотни и тысячи операций обращения к основной памяти. Таким образом, возникает идея организации процесса логического вывода, при котором потребуется как можно меньше запросов к внешнему источнику информации.

При стандартной организации обратного вывода после выбора отправной гипотезы машина осуществляет просмотр содержащих гипотезу заключений правил, переходя к предпосылкам и в свою очередь рекурсивно проверяя их истинность. Когда в данном процессе в качестве очередной гипотезы встречается начальный факт (он не присутствует в заключениях правил), для проверки его истинности необходимо обращение к базе данных или пользователю. При этом не все тестируемые факты в результате оказываются необходимыми для осуществления логического вывода, то есть порождения выбранной гипотезы.

Предлагаемый метод вывода на основе решения уравнений начинается с построения всех минимальных начальных прообразов в LP-структуре для атомов, соответствующих значениям объекта экспертизы. Далее в построенном множестве достаточно найти тот прообраз, который содержит лишь истинные факты, после чего сразу можно сделать заключение о соответствующем значении объекта экспертизы. Такой способ дает очевидное преимущество – исследуются лишь *минимальные* прообразы. Также предварительно можно исключить из процесса «противоречивые» прообразы, то есть содержащие одновременно альтернативные значения одного и того же объекта.

Однако существует более эффективный способ – приоритетный просмотр прообразов, содержащих значения наиболее «релевантных» объектов. Таковыми в частности могут считаться объекты, чьи значения присутствуют в максимальном количестве построенных прообразов. Тогда единственный отрицательный ответ пользователя на заданный вопрос исключает из рассмотрения сразу большое количество прообразов, что соответственно ускоряет исследование. Вторым признаком релевантности объекта – присутствие его значений в прообразах минимальной мощности. Таким образом, предпочтение отдается тем прообразам, проверка истинности которых потребует меньшего количества внешних запросов. Сочетая указанные два показателя релевантности, можно достичь результатов, по эффективности существенно превышающих возможности обычной машины вывода.

Переходя от экспертной системы к терминологии LP-структур, сформулируем соответствующую релевантному выводу следующую задачу нахождения истинного начального прообраза. Даны конечная атомно-порожденная решетка \mathbb{F} (атомы изображают элементарные факты) и на ней бинарное отношение R (представляет совокупность продукционных правил).

Не ограничивая общности, можно считать, что R является каноническим отношением, а правая часть уравнения (1) представляет собой конечное объединение атомов. Тогда достаточно рассмотреть уравнения с атомами в правой части:

$$R^L(X) = b \tag{4}$$

Пусть выбран атом $b \in \mathbb{F}$, которому соответствует общее решение уравнения (4) – множество $\{X\}$ всех его минимальных начальных прообразов. На множестве атомов решетки введем частично определенную булеву функцию $True$ (функцию «истинности»), которую можно доопределять путем обращения к внешнему источнику информации. В моделируемой продукционной системе интерпретация данной функции такова:

- $True(x) = 1$, если соответствующий факт x содержится в рабочей памяти;
- $True(x) = 0$, если достоверно известно, что x не может содержаться в рабочей памяти;
- $True(x) = null$, если проверка x еще не производилась.

Введем обозначения $T = \bigcup x_k (True(x_k) = 1)$; $F = \bigcup x_j (True(x_j) = 0)$.

Необходимо найти такой элемент $X^0 \in \{X\}$, что $X^0 \subseteq T$ (если он существует). В процессе решения задачи требуется также обойтись *как можно меньшим* количеством доопределений функции $True$.

Последнее требование лежит в основе метода *релевантного* вывода. Метод состоит из двух стадий: 1) решение уравнения – вычисление множества начальных прообразов и 2) нахождение среди них истинного прообраза. Выполнение указанных стадий может быть организовано в виде конвейера – независимо и параллельно. Работа на каждой из стадий также может быть распределена между параллельными вычислителями.

Однако при больших объемах баз знаний процесс построения *всех* минимальных прообразов может потребовать чрезмерного объема ресурсов компьютера. В связи с данным обстоятельством метод был модифицирован. *Кластерно-релевантный* LP-вывод предполагает построение кластеров начальных прообразов (подмножеств ограниченной мощности) с их динамическим релевантным исследованием.

Процессы решения сформулированной задачи представлены в работе в виде набора формальных алгоритмов. Возможно применение различных способов вычисления параметров релевантности, основные из которых: линейный, пропорциональный, бимаксимальный, вырожденный (с исключением одного из двух показателей).

Для повышения эффективности развиваемой методологии в диссертации разработаны также параллельные алгоритмы релевантного и кластерно-релевантного LP-вывода, а также параллельные алгоритмы вычисления истинных прообразов в LP-структурах. Заметим, что параллельным методам в продукционных системах было посвящено немало работ, к наиболее известным из которых относится фундаментальный труд А. Gupta¹¹. Эти методы могут быть использованы на более абстрактном уровне при компьютерной реализации LP-структур. Однако соответствующий раздел диссертационного исследования посвящен вопросам распараллеливания алгоритмов LP-вывода, стратегия которого направлена на снижение числа обращений к внешним устройствам. В

¹¹ Gupta A. Parallelism in Production Systems / A. Gupta. – Los Altos : Morgan Kaufmann, CA, 1987.

этом состоит основная особенность настоящей работы по отношению к другим параллельным реализациям продукционных систем, в которых минимизация по указанному параметру не применяется.

В результате использования методологии LP-вывода количество обращений к внешнему источнику информации существенно уменьшается и повышается эффективность их выполнения. Формальное обоснование данного утверждения представляет собой математически недоопределенную задачу, существенно зависящую от исходных данных. Далее в главе 4 эффективность LP-вывода будет обоснована результатами массивов экспериментов и их формальной статистической обработкой.

В **третьей главе** описана компьютерная реализация методов релевантного и кластерно-релевантного обратного вывода. Обсуждаются принципы кодирования решеток и LP-структур, используемые для эффективной программной реализации. Представлена архитектура объектно-ориентированного класса `ParallelLPStructure`, инкапсулирующего свойства и методы описанной в главах 1–2 теории LP-структур, включая нахождение логической редукции и решение продукционно-логических уравнений с параллельным вычислением прообразов.

Описывается созданная автором новая версия интегрированной среды разработки и эксплуатации продукционных систем `LPExpert`, использующая класс `ParallelLPStructure`. Основные расширения возможностей `LPExpert` таковы: выбор различных стратегий подсчета релевантности, параллельный LP-вывод, параллельное исследование прообразов, более гибкая и эффективная структура представления решеток с возможностью использования 64-разрядной архитектуры компьютера.

Четвертая глава содержит описание проведенных массивов экспериментов, в том числе со случайно генерируемыми базами знаний. Они призваны создать достаточный объем данных для исследования методами математической статистики. Таким образом формально обосновываются преимущества новой методологии обратного вывода в различных ее модификациях.

При выполнении алгоритмов LP-вывода, описанных в главе 2, может производиться значительно больше вычислений по сравнению с обычным обратным выводом. Применяемая в настоящей работе стратегия направлена на уменьшение количества обращений за информацией к *внешнему источнику*, а также и общего времени работы за счет *распараллеливания задач*. Как уже говорилось, формальное обоснование уменьшения числа запросов представляет математически недоопределенную задачу, существенно зависящую от исходных данных. С учетом данного обстоятельства приобретает особое значение формальное сравнительное исследование результатов проводимых экспериментов. Аппаратом такого исследования может служить математическая статистика.

В разделе 4.1 представлены основная терминология и последовательность применения некоторых статистических методов обработки данных, а именно:

статистическая проверка статистических гипотез, сравнение двух дисперсий нормальных генеральных совокупностей, сравнение двух средних значений нормальных генеральных совокупностей. Рассматриваются возможности исследования в пакете Statistica: Т-критерий, F-критерий, U-тест Манна-Уитни, дисперсионный анализ.

Далее в главе приводятся результаты использования таких методов. Сравниваются количества внешних запросов при использовании обычного обратного вывода, релевантного вывода и его модификаций, а также кластерно-релевантного вывода.

С целью обоснования эффективности предложенных в работе алгоритмов LP-вывода было проведено порядка 500 автоматизированных тестов с «формальными» базами знаний, генерируемыми случайным образом с возможностью контроля «глубины», количества генерируемых правил и объектов. Описанные эксперименты позволяют определить количество выполняемых запросов к внешнему источнику информации в зависимости от объема исходных фактов и правил. Показано, что число внешних запросов в среднем снижается на 15–20%.

Кроме этого, представлены итоги сравнения времени выполнения релевантного вывода с использованием параллельных вычислений и без них. Все полученные результаты сопровождаются иллюстрациями, сгенерированными при обработке данных в пакете Statistica 6, необходимыми диаграммами и графиками, по которым также можно судить о преимуществах рассматриваемых алгоритмов.

В завершение главы полученные при статистических исследованиях результаты систематизируются. Они приводят к следующим итогам.

- Релевантный вывод сокращает число запросов к внешнему источнику в среднем на 15–20% по сравнению с обычным обратным выводом.
- Модификация релевантного вывода с пропорциональным подсчетом релевантности также обеспечивает прирост в эффективности, заключающийся в дальнейшем уменьшении числа запросов.
- Кластерно-релевантный вывод дает хорошие результаты по сравнению с обычным обратным выводом в случае работы с базами данных больших размеров, для которых релевантный вывод не может быть применен из-за ограниченности ресурсов компьютера.
- Параллельная реализация релевантного LP-вывода позволяет сократить время выполнения рассматриваемых процессов в среднем на 20%. Кроме того, по полученным графикам установлено оптимальное количество потоков, дающее наибольший прирост к эффективности для текущей конфигурации компьютера.

Сделанные выводы наглядно подтверждаются построенными графиками и диаграммами.

В **Заключении** подводятся итоги и обсуждаются некоторые направления дальнейших исследований.

Приложение А диссертации содержит таблицы результатов тестирования методологии LP-вывода, проведенного в рамках пакета программ LPExpert.

В **Приложении В** представлены исходные тексты некоторых основных модулей программной системы, разработанных лично автором.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ ДИССЕРТАЦИИ И ПЕРСПЕКТИВЫ

Настоящая диссертационная работа представляет собой исследование в области формальных знаний – логических систем продукционного типа, составляющих основу многих моделей в информатике. Его основная задача – разработка и изучение метода релевантного LP-вывода, с целью развития его идей до работающей на практике новой технологии ускорения логического вывода и верификации знаний, а также обоснование преимуществ этой технологии в сравнении с другими подходами к аналогичным задачам. В рамках проведенного исследования получены следующие основные результаты.

- Сформулирована и исследована обобщенная модель LP-структуры, расширяющая область применения теории LP-структур.

- Усовершенствована схема процесса решения продукционно-логического уравнения. В результате из этого процесса исключен избыточный этап, связанный с идентификацией приближенных решений.

- Разработан новый метод обратного вывода в логических системах продукционного типа, обеспечивающий снижение числа обращений к внешним устройствам.

- Предложены и апробированы различные способы выбора параметров релевантности для процессов релевантного и кластерно-релевантного LP-вывода, предоставляющие гибкие возможности управления обратным выводом.

- Разработаны параллельные алгоритмы релевантного и кластерно-релевантного LP-вывода, а также параллельные алгоритмы вычисления истинных прообразов в LP-структурах, повышающие быстродействие LP-вывода.

- Создана новая версия интегрированного программного пакета LPExpert, предназначенного для разработки и эксплуатации продукционно-логических систем. Она реализует все перечисленные выше алгоритмы, параллельные вычисления и стратегии релевантности, а также поддерживает 64-разрядную архитектуру процессора.

- Выполнены массивные вычислительные эксперименты и проведено подробное исследование полученных результатов методами математической статистики. В результате формально обоснованы преимущества нового метода обратного вывода в различных его модификациях.

Дальнейшие исследования на рассматриваемом направлении могут быть связаны с выбором LP-структур более сложных типов и соответствующих им предметных областей, перенос концепций настоящей работы на эти модели. Например, при использовании в качестве решетки алгебры Линденбаума–Тарского моделируемые условные правила смогут в качестве предпосылок и

заклучений содержать формулы пропозиционального исчисления. Таким образом можно будет рассматривать расширенные модели логических систем и решать для них задачи ускорения и распараллеливания процессов обратного вывода с общей стратегией снижения запросов к внешним источникам информации. Общие методы исследования при этом останутся прежними.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

1. *Болотова С.Ю.* Алгоритмы релевантного обратного вывода, основанные на решении продукционно-логических уравнений / С.Ю. Болотова, С.Д. Махортов // Искусственный интеллект и принятие решений. – 2011. – № 2. – С. 40–50.
В работе [1] Болотовой С.Ю. принадлежат математические результаты и алгоритмы.
2. *Болотова С.Ю.* Алгебраическая модель релевантного обратного вывода на основе решения уравнений / С.Ю. Болотова // Математическое моделирование. – 2012. – № 12, т. 24. – С. 3–8.
3. *Болотова С.Ю.* Применение многопоточности в релевантном LP-выводе / С.Ю. Болотова // Нейрокомпьютеры. Разработка, применение. – 2013, № 9. – С. 53–57.
4. *Болотова С.Ю.* Реализация многопоточности в релевантном LP-выводе / С.Ю. Болотова // Программная инженерия. – 2014, № 1. – С. 12–18.
5. *Болотова С.Ю.* О релевантном обратном выводе в системах знаний продукционного типа / С.Ю. Болотова, С.Д. Махортов // Третья Всероссийская конференция с международным участием «Знания – Онтологии – Теории» (ЗОНТ-2011). Новосибирск, 3–5 октября 2011г.: Материалы конференции. – Новосибирск: Институт математики им. С.Л. Соболева СО РАН, 2011. – Т.1. – С. 73–77.
В работе [5] Болотовой С.Ю. принадлежат математические результаты и алгоритмы.
6. *Болотова С.Ю.* Релевантный обратный вывод и верификация знаний на основе решения уравнений / С.Ю. Болотова // X Всероссийская научная конференция «Нейрокомпьютеры и их применение» (НКП-2012), Москва, 20 марта 2012 года. Тезисы докладов. – М: МГППУ, 2012. – С. 12.
7. *Болотова С.Ю.* Алгебраическая модель релевантного обратного вывода на основе решения уравнений / С.Ю. Болотова // Современные проблемы прикладной математики и информатики (МРАМС'2012): Тезисы докладов международной молодежной конференции-школы (Дубна, 22–27 августа 2012г.). – Дубна: ОИЯИ, 2012. – С. 67–69.
8. *Болотова С.Ю.* Использование параллельных вычислений в методе релевантного обратного вывода / С.Ю. Болотова // XI Всероссийская научная конференция «Нейрокомпьютеры и их применение» (НКП-2013), Москва, 19 марта 2013 года. Тезисы докладов. – М: МГППУ, 2013. – С. 16.

9. *Bolotova S.* Multi-threaded relevant LP-inference / S. Bolotova, S. Makhortov // Distributed Intelligent Systems and Technologies (DIST'2013): Proceedings of the International Conference (St. Petersburg, July 1–4, 2013). – St. Petersburg: Spbstu, 2013. – Pp. 7–14. (In English)
В работе [9] Болотовой С.Ю. принадлежат математические результаты и алгоритмы.
10. *Bolotova S.* Using multi-threading in the relevant LP-inference method / S. Bolotova, S. Makhortov // Mathematical Modeling and Computational Physics (MMCP'2013): Book of Abstracts of the International Conference (Dubna, July 8–12, 2013). – Dubna: JINR, 2013. – P. 58. (In English)
В работе [10] Болотовой С.Ю. принадлежат математические результаты и алгоритмы.
11. *Болотова С.Ю.* Использование параллельных вычислений в методе релевантного обратного вывода / С.Ю. Болотова // Материалы Всероссийской конференции с международным участием «Знания – Онтологии – Теории» (ЗОНТ-2013, Новосибирск, 8–10 октября 2013 г.). – Новосибирск: Институт математики им. С.Л. Соболева СО РАН, 2013. – Т.1. – С. 50–59.
12. *Болотова С.Ю.* Стратегии релевантного обратного вывода на основе решения логических уравнений / С.Ю. Болотова // Современные проблемы прикладной математики, теории управления и математического моделирования (ПМТУММ-2011): материалы IV Международной научной конференции (Воронеж, 12–17 сентября 2011 г.). – Воронеж : ИПЦ ВГУ, 2011. – С. 37–40.
13. *Болотова С.Ю.* Статистика релевантного обратного вывода / С.Ю. Болотова // Современные проблемы прикладной математики, теории управления и математического моделирования (ПМТУММ-2012): материалы V Международной научной конференции (Воронеж, 11–16 сентября 2012 г.). – Воронеж : ИПЦ ВГУ, 2012. – С. 48–50.
14. *Болотова С.Ю.* Статистические результаты релевантного обратного вывода / С.Ю. Болотова // Актуальные проблемы прикладной математики, информатики и механики: Сб. трудов Международной конференции (Воронеж, 26–28 ноября 2012 года), ч.2. – Воронеж : ИПЦ ВГУ, 2012. – С. 45–49.
15. *Болотова С.Ю.* Библиотека ParallelLPInference / С.Ю. Болотова // Свидетельство о государственной регистрации программы для ЭВМ. – М.: Федеральная служба по интеллектуальной собственности. – № 2013617293 от 04.10.2013.