на правах рукописи

ЗАБЕЖАЙЛО МИХАИЛ ИВАНОВИЧ

КОМБИНАТОРНЫЕ СРЕДСТВА ФОРМАЛИЗАЦИИ ЭМПИРИЧЕСКОЙ ИНДУКЦИИ

05.13.17 – теоретические основы информатики

Автореферат
Диссертации на соискание ученой степени
доктора физико-математических наук

Москва – 2016

Диссертационная работа выполнена в Лаборатории Интеллектуального анализа данных и автоматизированной поддержки научных исследований Федерального государственного учреждения «Федеральный исследовательский центр «Информатика и управление» Российской академии наук»

Научный консультант	Доктор технических наук, профессор, Заслуженный деятель науки РФ ФИНН Виктор Константинович
Официальные оппоненты	Чл.-корр. РАН, д.т.н., профессор ЖЕЛТОВ Сергей Юрьевич, Генеральный директор ФГУП «Государственный научно-исследовательский институт авиационных систем».
	Д.ф-м.н., РЕДЬКО Владимир Георгиевич, заместитель руководителя Центра оптико-нейронных технологий ФГУ «Федеральный научный центр Научно-исследовательский институт системных исследований Российской академии наук»
	Д.ф-м.н., профессор БЕНИАМИНОВ Евгений Михайлович, Институт лингвистики Российского Государственного Гуманитарного Университета, заведующий Кафедрой математики, логики и интеллектуальных систем в гуманитарной сфере
Ведущая организация	Федеральное государственное автономное образовательное учреждение высшего профессионального образования «Московский физико-технический институт (государственный университет)».

Защита состоится *15 декабря 2016* г. в 13:00 на заседании диссертационного совета Д 002.073.05 при Федеральном исследовательском центре «Информатика и управление» Российской академии наук (ФИЦ ИУ РАН) по адресу: 119333, г.Москва, ул.Вавилова, д. 40.

С диссертацией можно ознакомиться в библиотеке ФИЦ ИУ РАН а также на сайте http://www.frccsc.ru/

Автореферат разослан ____ _____ 2016 года.

Ученый секретарь
диссертационного совета Д 002.073.05
д.ф-м.н., профессор В.В.Рязанов

Актуальность темы исследования. Диссертационная работа посвящена проблемам разработки математических моделей, методов и алгоритмов извлечения зависимостей из коллекций исходно заданных прецедентов – примеров и контрпримеров.

Проблема эмпирической индукции, сформулированная, в частности, как задача обучения на прецедентах, хорошо известна специалистам, характеризуется разнообразием методов ее решения и обширной литературой[1]. Однако, большинство имеющихся публикаций по этой тематике ориентированы на обработку числовых данных. Отдельной и сравнительно мало изученной областью является проблематика извлечения зависимостей из данных (описаний прецедентов), представленных *нечисловыми* объектами (множествами сущностей и отношениями на них – графами, цепочками символов конечного алфавита и т.п. реляционными структурами) и характеризуемых дополнительно числовыми значениями ряда существенных параметров. Поиск (извлечение из исходно заданных выборок прецедентов) зависимостей вида **ОБЪЕКТ=>СВОЙСТВА** на нечисловых объектах, характеризуемых в том числе и числовыми параметрами, представляет собою важную и актуальную область как теоретических, так и прикладных исследований в рамках интеллектуального анализа данных. Ее особый статус как самостоятельной сферы исследований в области искусственного интеллекта (так называемой проблематики *phenomenal data mining*) начал активно изучать *J.McCarthy*. Помимо нечисловой (неметрической) природы анализируемых зависимостей, весьма важной здесь оказывается возможность оперировать в том числе и с *малыми* (статистически незначимыми) *выборками* прецедентов (примеров–объектов, которые наделены целевыми свойствами, и контрпримеров–объектов, которые не обладают целевыми свойствами), причем оперировать так, чтобы иметь достаточные основания для принятия порождаемых результатов (извлекаемых из данных эмпири-

[1] Так, например, только в монографии *Кайберг Г. Вероятность и индуктивная логика***.** *- М.: Прогресс, 1978. - 374 С*. список литературы содержит более 1000 наименований.

ческих зависимостей). Не менее существенными оказываются также возможности выделять совокупности факторов, определяющих наличие (или же отсутствие) изучаемых свойств у анализируемых прецедентов, - структурные «носители» («*причины*» проявления) этих свойств. Наконец, учитывая, что при организации машинного обучения на нечисловых объекта, как правило, приходится сталкиваться с комбинаторно сложными проблемами (в частности – NP-полными, перечислительно полными и т.п. задачами), следует обратить особое внимание на характеристики *вычислительной сложности* разрабатываемых моделей, методов и алгоритмов извлечения эмпирических зависимостей из структурных данных, рассмотреть возможности *оптимизации* необходимого для поиска решений *комбинаторного перебора* вариантов. Представленный комплекс требований определяет актуальную задачу теоретической информатики, вариант решения которой и предлагается в данной диссертационной работе.

В области приложений математические модели подобного сорта а также построенные на их основе программные системы интеллектуального анализа данных весьма актуальны, в частности, в задачах компьютерного прогнозирования физиологических активностей химических соединений (как – лекарственных, так и, наоборот, представляющих угрозу для жизни - токсичности, канцерогенности и т.п. - свойств). Анализ причинности актуален, например, для задач медицинской и технической диагностики (где выделение причинных факторов позволяет организовать эффективное противодействие их негативному влиянию). Наконец, возможности оптимизировать объемы вычислений при порождении зависимостей из эмпирических данных весьма актуальны для проблематики так называемых Big Data.

Центральным объектом исследования является сформированная в диссертационной работе процедурная схема порождения зависимостей на прецедентах *нечислового характера*. Демонстрируется (в том числе – путем доказательства ряда строгих формальных утверждений) корректность такой схемы, построены оценки сложности вычислений входящих в нее комбинаторных задач.

Предложены эффективные механизмы оптимизации комбинаторного перебора вариантов при поиске эмпирических зависимостей (метод *последовательных приближений*, реализуемый в том числе и в режиме *параллельных* вычислений).

Основной задачей диссертационного исследования является создание инструментария интеллектуального анализа данных для эффективного решения задач обучения на прецедентах, которые описаны как наделенные внутренней структурой *нечисловые* объекты (дополненные в ряде случаев числовыми значениями параметров, релевантных свойствам этих объектов).

Одним из наиболее известных и весьма детально развитых направлений в области разработки математических моделей и методов *машинного обучения* на прецедентах является проблематика так называемой *Теории статистического обучения* (*statistical learning theory – SLT)* - см., например, как уже ставшие классическими работы *В.Н.Вапника* и *А.Я.Червоненкиса*, так и ряд новейших достижений, в частности, результаты *К.В.Воронцова* по теории надежности обучения на прецедентах и др.. Ключевой проблемой *Теории статистического обучения* является формирование оценок вероятности ошибки при переносе (экстраполяции) построенных на обучающей выборке зависимостей на новые (не входящие в эту выборку) объекты. Однако в случае малых (статистически не значимых) выборок здесь на сегодняшний день имеется целый ряд еще не решенных вопросов.

Не менее известен подход, опирающийся на математическую технику так называемых *корректных алгебр* над множеством некорректных (эвристических) алгоритмов, предложенную *Ю.И.Журавлевым* и успешно развиваемую его школой. Алгебраический подход к проблеме синтеза корректных алгоритмов открыл принципиально новые возможности при решении различных классов плохо формализованных задач.

В более широком контексте (т.е. в ситуации, когда для порождения зависимостей из данных используются не только метрические модели и алгоритмы – см., например, работы *D.Michie, T.Mitchell, C.M.Bishop, P.Langley, M.Mohri* и

др.) весьма чувствительны ограничения на область практического применения подобного инструментария, которые связаны с, как правило, экспоненциально быстро растущими оценками сложности вычислений, характерными для возникающих здесь комбинаторных задач.

Достаточно широко распространенными технологиями решения обсуждаемой нами задачи обучения на прецедентах являются также так называемые

- *нейронные сети* (см., например, работы *Т.Кохонена, M.Chester, В.В.Круглова, Д.Рутковской* и др.), где при поиске решения используется специальный вариант математических моделей многопараметрической нелинейной оптимизации, и
- *генетические алгоритмы* (см. работы *Л.А.Гладкова, Д.Рутковской, Л.Шашкина* и др.), где решение задач оптимизации при выборе наиболее подходящего класса зависимостей ищется моделированием локальных процедур случайного выбора, вариации и\или комбинирования анализируемых параметров по аналогии с механизмами естественного отбора в живой природе.

Однако, в анализе связей вида **ОБЪЕКТ => СВОЙСТВА** при поиске ответа на вопрос «*почему возникает анализируемое явление*?», предлагаемый нейронными сетями и генетическими алгоритмами вариант ответа вида «*такой результат дал нам используемый алгоритм расчета*» вряд ли сможет убедить независимого эксперта в наличии достаточных оснований для принятия полученных соответствующим инструментарием результатов.

Итак, целый ряд вопросов, характерных для рассматриваемой проблематики обучения на прецедентах продолжительное время остается открытым в части операций с объектами нечислового характера. Примером реализации успешного подхода к решению задач этого типа является так называемый ДСМ-метод автоматического восстановления зависимостей из эмпирических данных, предложенный и успешно развиваемый *В.К.Финном* и его школой. Тем не менее, и для используемых до настоящего времени программных реализаций ДСМ-метода в приложениях весьма критичными являются практические ограничения на размеры реально обрабатываемых в приемлемое время

выборок прецедентов, обусловленные вычислительной сложностью исполняемого в рамках ДСМ-метода комбинаторного перебора вариантов.

Целью диссертационной работы является разработка математического аппарата (методов, моделей и алгоритмов) для решения задач извлечения зависимостей из эмпирических данных (описаний прецедентов) - *нечисловых* объектов сложной структуры (характеризуемых также числовыми значениями существенных параметров), и демонстрация возможностей эффективного применения этого аппарата в приложениях из различных предметных областей.

Научная новизна. В процессе разработки предложенных в диссертационной работе математических методов, моделей и алгоритмов

- для идентификации порождаемых эмпирических зависимостей развита оригинальная техника формирования классов эквивалентности на исходно задаваемом множестве прецедентов. Такие классы эквивалентности реконструируются по классам сходства прецедентов, которые формируются на основе отношения структурного сходства описаний прецедентов, формализуемого с помощью соответствующей алгебраической операции;
- для анализа корректности (выполнимости условий достаточности оснований для принятия) порождаемых эмпирических зависимостей а также прогноза их расширения (экстраполируемости) на описания новых прецедентов предложены и используются (в соответствующей процедурной схеме) особые комбинаторные объекты – диаграммы частичного порядка взаимной вложимости построенных классов эквивалентности;
- для нескольких типов структурных описаний исходно заданных прецедентов (множеств, графов, цепочек символов конечного алфавита, векторов числовых значений параметров и др.) получены оценки вычислительной сложности, характерные для переборных задач, которые возникают при формировании подобных классов эквивалентности и диаграмм их вложимости;
- развита оригинальная алгоритмическая техника формирования приближенных описаний диаграмм частичного порядка анализируемых классов эквивалентности (так называемая процедурная конструкция *приближенного ДСМ-*

метода), которая позволяет организовать целенаправленный управляемый перебора вариантов при порождении эмпирических зависимостей рассматриваемого типа, в том числе:

a) строить в первую очередь (используя так называемые каркасы псевдо-деревьев замыканий Галуа, формируемых на множестве исходно заданных прецедентов, причем - строить полиномиально быстро) так называемые *полезные* (для прогноза свойств новых объектов, для проверки выполнимости условия каузальной полноты и др.) эмпирические зависимости,

а затем, если потребуется,

b) достраивать (уже, вообще говоря, экспоненциально сложными вычислениями) множество этих зависимостей до состояния, в котором восстановлены *все* содержащиеся в исходных данных эмпирические зависимости;

- сформулированы и доказаны утверждения, определяющие корректность предложенной процедурной конструкции приближенного ДСМ-метода;
- представлен вариант реализации приближенного ДСМ-метода в режиме *параллельных вычислений*.

Работоспособность предложенных математических моделей и алгоритмов продемонстрирована на примерах решения конкретных прикладных задач.

Методы исследования. В диссертационной работе использованы:

- логико-комбинаторные и алгебраические методы дискретной математики;
- методы взаимной сводимости и построения оценок вычислительной сложности трудных переборных задач,
- методы анализа и дискретной оптимизации вычислительных алгоритмов.

Теоретическая значимость. Разработанные математические модели, методы и алгоритмы позволяют организовать порождение эмпирических зависимостей в том числе на малых выборках сложно структурированных описаний прецедентов, исходно заданных для обучения.

Предложенная в диссертационной работе математическая техника формирования и анализа диаграмм взаимной вложимости классов эквивалентности прецедентов демонстрирует комбинаторную природу хорошо известного

в исследованиях по искусственному интеллекту и ассоциируемого с проблемой эмпирической индукции класса эвристик. (Содержательные свойства этих эвристик подробно рассматривались *Д.С.Миллем*, *Д.Пойа*, *Ч.С.Пирсом* и другими. Подходы к формализации предложены *В.К.Финном* и его школой).

Развитая процедурная конструкция целенаправленного построения приближенных описаний таких диаграмм формирует основу для создания специального класса программных систем искусственного интеллекта, ориентированных на использование формализованных моделей правдоподобных рассуждений (построения индуктивных обобщений, рассуждений по аналогии, формирования абдуктивных объяснений и т.п.).

Практическая значимость. Алгоритмическая техника формирования приближенных описаний диаграмм частичного порядка рассматриваемых классов эквивалентности (так называемая процедурная конструкция приближенного ДСМ-метода) дает возможности оперировать при восстановлении эмпирических зависимостей рассматриваемого класса исходными выборками данных любого (в т.ч. – большого) размера.

Предложенная техника формирования структурных описаний диаграмм частичного порядка классов эквивалентности позволяет организовать *параллельную обработку* данных (в том числе, для промышленных приложений - на базе масштабируемых облачных вычислений).

Предложенная техника восстановления зависимостей успешно применяется при решении ряда прикладных задач интеллектуального анализа данных.

Положения, выносимые на защиту

1. Предложена *алгебраическая формализация* для используемой (построенной на базе комплекса эвристик эмпирической индукции) процедурной конструкции *извлечения зависимостей* из структурных описаний исходно заданных для обучения прецедентов, которая обеспечивает унифицированные возможности оперировать *различными типами данных*.
2. Идентифицирован и исследован *специальный класс комбинаторных объектов* (диаграмм взаимной вложимости классов эквивалентности различных

типов описаний объектов-прецедентов), *определяющих особенности алгоритмики* порождения эмпирических зависимостей предлагаемым способом.

3. Процедурная конструкция ДСМ-метода расширена на *новые типы данных* (структурные описания *нечисловых объектов*, дополненные *числовыми значениями* существенных параметров).
4. Предложено доказательство *оценок сложности вычислений*, требуемых для решения основных переборных задач, которые возникают при формализации эмпирической индукции в процессе порождения диаграмм вложимости классов эквивалентности при обработке *различных типов данных* (вариантов структурных описаний исходно заданных прецедентов).
5. Предложен метод *дискретной оптимизации* и *управления* комбинаторным *перебором* вариантов при формировании рассматриваемых диаграмм и анализе переносимости порождаемых эмпирических зависимостей на описания новых - ранее еще не изученных - прецедентов (*метод последовательных приближений* при формировании диаграмм вложимости, реализуемый в том числе и в режиме *параллельных вычислений*, что в совокупности позволяет *снять ограничения* на *размеры* эффективно обрабатываемых *исходных коллекций прецедентов*).
6. Разработано математическое *обоснование корректности* процедурной схемы *метода последовательных приближений*, предложенного для формирования диаграмм вложимости классов эквивалентности прецедентов.
7. *Практическая значимость и эффективность* предложенных математических моделей, методов и алгоритмов подтверждена результатами, полученными при решении прикладных задач в различных предметных областях (при решении прикладных задач интеллектуального анализа данных средствами ДСМ-метода автоматического порождения зависимостей).

Область исследования, согласно Паспорту специальности 05.13.17 – «Теоретические основы информатики»:

- *разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных* (п.5);

- *моделирование формирования эмпирического знания* (п.7);
- *исследование и когнитивное моделирование интеллекта, включая моделирование поведения, моделирование рассуждений различных типов* (п.8);

а также

- *исследование информационных структур, разработка и анализ моделей информационных … структур* (п.2);
- *исследование и разработка средств представления знаний* (п.4).

В соответствии с обозначенными в формуле специальности 05.13.17 – «Теоретические основы информатики» направлениями, область данного исследования характеризуют в том числе и:

«…*исследования методов преобразования информации в данные и знания*; … *исследование … моделей данных и знаний*, … *методов машинного обучения и обнаружения новых знаний*; *исследования принципов создания и функционирования … программных средств автоматизации указанных процессов*».

Таким образом, выполненное в диссертационной работе исследование комбинаторной структуры соответствующего класса формальных моделей эмпирической индукции соответствует специальности 05.13.17 – «Теоретические основы информатики».

Достоверность и обоснованность. Апробация работы.

Достоверность и обоснованность полученных результатов, научных положений и выводов подтверждается строгостью и корректностью использования комплекса методов математической логики, алгебры, теории сложности вычислений и комбинаторной оптимизации, строгостью и корректностью приведенных в диссертационной работе математических утверждений и их доказательств. Выполнена экспериментальная проверка полученных результатов в ряде задач интеллектуального анализа данных средствами ДСМ-метода автоматического порождения гипотез, в рамках которого реализован синтез комплекса эвристик (индукции, аналогии и абдукции). Проведённые практические исследования коллекций эмпирических данных из различных прикладных предметных областей подтверждают обоснованность и практическую полезность представленных в диссертационной работе положений и выводов.

Основные положения и результаты работы докладывались и обсуждались на всероссийских и международных конференциях, конгрессах, чтениях и семинарах. В том числе – на Международных конференциях НТИ (ВИНИТИ РАН), Национальных конференциях с международным участием «Искусственный интеллект» (Российская ассоциация искусственного интеллекта), Конференциях и семинарах IEEE (США), Российско-британской конференции «Идеи Д.С. Милля об индукции и логике наук о человеке и обществе в когнитивных исследованиях и системах искусственного интеллекта» (РГГУ), научных семинарах ВИНИТИ РАН, МФТИ, ВМК МГУ, ВЦ РАН, ИПУ РАН и др..

Материалы данной диссертационной работы легли в основу спецкурса «Формальные модели рассуждений в задачах компьютерного восстановления зависимостей из данных», читавшегося автором на протяжении ряда лет студентам старших курсов на Факультете управления и прикладной математики Московского физико-технического института.

Публикации

По теме диссертации опубликовано 37 работ[2] (см. Список публикаций), в том числе 17 - в изданиях из списка ВАК и 2 - в изданиях IEEE.

Личный вклад автора. В диссертационной работе представлены только результаты, которые полученные лично автором: исследование проблематики восстановления зависимостей по прецедентам, формулировки и доказательства утверждений, постановки задач, методы и алгоритмы их решения. Из совместных публикаций в диссертацию включены лишь результаты автора.

Структура и объем работы. Диссертационная работа состоит из Введения, 6 Глав, Заключения, Списка литературы (525 наименований на 33 страницах) и 8 Приложений. В работе - 17 рисунков и 19 таблиц). Общий объём работы - 379 стр. (280 стр.- основной текст и 99 стр. – Приложения). Основные математические результаты работы представлены в **Разделах 2.1** и **3.1**, **Главах 4** и **5**. Сформулированы и доказаны более 100 строгих утверждений (теорем, лемм и

[2] Без учета переводов и перепечаток.

др.), которые вместе с набором соответствующих алгоритмов характеризуют различные математические аспекты формализации комплекса эвристик эмпирической индукции. Содержательные свойства используемых эвристик анализируются в **Разделах 2.2**, **3.2** и **6.4**. Работоспособность развиваемого подхода при решении прикладных задач продемонстрирована в **Приложениях**.

Содержание работы.

Глава 1 носит вводный характер и обсуждает общие характеристики области интеллектуального анализа данных (ИАД) и места, отводимого в ней проблематике математических моделей и методов восстановления зависимостей из накапливаемой эмпирической информации. Задача формализации эмпирической индукции (ЭИ) – *обучения на прецедентах* – рассматривается в первую очередь на примере так называемого ДСМ-метода автоматического порождения эмпирических зависимостей из данных и формулируется в контексте исследований в области искусственного интеллекта – как задача адекватной (позволяющей контролировать доказуемую корректность их использования) формализации ассоциируемого с понятием ЭИ комплекса эвристик (индуктивного обучения, рассуждений по аналогии и абдуктивных объяснений).

Все основные задачи данного диссертационного исследования естественным образом вытекают из представленного в этой **Главе** обзора проблем, характерных для анализируемой области ИАД (в ее текущем состоянии).

В качестве основной области исследований выбран важный (и мало изученный на текущий момент) частный случай общей проблематики обучения на прецедентах – работа с данными нечислового характера. Рассмотрены особенности анализа причинно-следственных связей между описаниями нечисловых объектов (реляционных структур, возможно, дополненных числовыми значениями некоторых существенных параметров) и множествами присущих им свойств. Особо подчеркнута актуальность развития математических моделей и методов, позволяющих получать корректные решения таких задач на малых (статистически незначимых) выборках прецедентов. Обозначены актуальные для современного уровня развития ИАД проблемы и ограничения.

Приведены примеры актуальных областей приложения рассматриваемых технологий ИАД (в частности - задачи компьютерного прогнозирования физиологических активностей химических соединений, автоматизированной обработки больших массивов полнотекстовых документов и др.).

Даны общие представления об архитектуре интеллектуальных компьютерных систем, используемых в анализе данных. Обсуждаются возможности и ограничения наиболее продвинутых математических моделей и техник компьютерного анализа данных. Сформулированы характеристики исследовательской ситуации, в которой адекватным представляется использование именно интеллектуальных систем анализа данных.

В **Главе 2** описываются базовые смысловые элементы организации интеллектуального анализа данных средствами ДСМ-метода. Задача обучения на прецедентах рассматривается в следующей общей формулировке:

Даны: **1**) множество (примеров) *объектов*, обладающих целевыми *свойствами*, и множество[3] (контрпримеров) объектов, не обладающих целевыми свойствами,

2) новый объект (или же некоторым явным образом представленное множество таких объектов)

Требуется: оценить *наличие* (или отсутствие) *целевых свойств* у нового *объекта* (новых объектов из заданного множества), т.е.

3) дать соответствующий *прогноз* (о наличии целевых свойств) и

4) предъявить достаточные *основания* (неоспоримые *аргументы*), которые позволяют *принять* этот прогноз.

Вопрос о *достаточности оснований* для принятия результатов выполняемой процедуры прогнозирования может быть уточнен, в частности, до формулировки следующего вида:

> *Имеются ли возможности при решении рассматриваемой задачи обучения на прецедентах сформировать средства контроля надежности порождаемых результатов, сопоставимые, например, со средствами, которые обеспечивают надежность результатов достоверного (дедуктивного логического) вывода?*

[3] В исходном состоянии, - возможно, пустое.

Дано подробное описание общей для различных типов нечисловых данных (дополнительно характеризуемых также и числовыми параметрами) процедурной схемы, предлагаемой для решения поставленной задачи. Рассматриваемая схема основана на формализации (алгебраическими и логико-математическими средствами) трех базовых эвристик – индуктивного обучения, выводов по аналогии и абдуктивного объяснения – и определяет строгие условия корректности применения этих эвристик.

В **Разделе 2.1** предложена схема формализации задачи обучения на прецедентах, выполненной алгебраическими средствами. В основе этой конструкции – анализ структурного сходства описаний прецедентов (примеров, характеризуемых наличием целевых свойств, и контрпримеров, характеризуемых их отсутствием), в котором понятие собственно сходства формализуется как бинарная алгебраическая операция. Каждая такая операция $\otimes$ для всех элементов множества $\mathbf{\Omega}$ исходно заданных прецедентов обязана удовлетворять трем классическим условиям (порождать из множества $\mathbf{\Omega}$ полурешетку):

$a \otimes a = a$
$a \otimes b = b \otimes a$
$(a \otimes b) \otimes c = a \otimes (b \otimes c) = a \otimes b \otimes c.$

На базе операции сходства $\otimes$ определяется отношение сходства $\mathbf{R}^{\otimes}$, которое характеризуемое непустыми результатами вычисления этой операции:

a $\mathbf{R}^{\otimes}$b (т.е. $\langle a,b \rangle \in \mathbf{R}^{\otimes}$) имеет место тогда и только тогда, когда

$a \otimes b \neq \llcorner \varnothing \lrcorner$ (где $\llcorner \varnothing \lrcorner$ есть пустой объект того же типа, что a и b).

С помощью отношения $\mathbf{R}^{\otimes}$ на множестве описаний исходно заданных прецедентов «одного знака»[4] строятся классы сходства вида:

$$\mathbf{T}(a) = \{b \mid a\ \mathbf{R}^{\otimes} b\ \}.$$

Для корректного порождения полурешетки сходств прецедентов исходного множества $\mathbf{\Omega}$ формируется множество *Dom*($\mathbf{\Omega}$): сперва в *Dom*($\mathbf{\Omega}$) помещаются все элементы $\mathbf{\Omega}$, затем в *Dom*($\mathbf{\Omega}$) для каждой пары уже существующих в нем

[4] Т.е. отдельно – на примерах и контрпримерах.

элементов A и B добавляется и результат операции $A \otimes B$, но при условии, что он не является пустым:

$$\{[A \in Dom(\mathbf{\Omega})] \,\&\, [B \in Dom(\mathbf{\Omega})] \,\&\, [(A \otimes B) \neq \ulcorner \varnothing \urcorner]\} \rightarrow [\,(A \otimes B) \in Dom(\mathbf{\Omega})\,].$$

Затем фиксацией каждого конкретного результата $V = V_0$ вычисления сходства $\otimes$ на элементах множества $Dom(\mathbf{\Omega})$

$$A \otimes B = V = V_0$$

выделяются соответствующие подклассы сформированных классов сходства:

$$\mathbf{E}_{V_0}(a) = \{b \mid a \otimes b \otimes V_0 = V_0\}.$$

Показано, что справедливо:

Утверждение 2.1.1

Каждый подкласс $\mathbf{E}_{V_0}(a)$ как подмножество исходно заданного множества прецедентов $\mathbf{\Omega}$ есть класс эквивалентности.

□

Также показано (***Следствие 2.1.2***) что пространство толерантности $\langle \mathbf{\Omega}, \mathbf{R}^{\otimes} \rangle$ может быть представлено в виде объединения всех порождаемых указанным выше способом классов эквивалентности $\mathbf{E}_{V_0}(\Omega, \otimes)$. При этом каждый подобный класс эквивалентности может быть порожден специальным оператором $\boldsymbol{Clos}^{\otimes}: 2^{\Omega} \rightarrow 2^{\Omega}$, реализующим один и тот же набор действий при работе с различными типами описаний прецедентов:

> *имея объекты* a,b, … , c *из* $\mathbf{\Omega}$ *а также их сходство* v= ((a⊗b)⊗… ⊗c) *найти все оставшиеся в* $\mathbf{\Omega}$ *объекты* d_i, *сходство которых с имеющимся* v *есть именно этот объект* v.

Показано, что этот оператор $\boldsymbol{Clos}^{\otimes}$ для любых $\boldsymbol{O}$, $\boldsymbol{O_i}$ и $\boldsymbol{O_j}$ из 2^{Ω} удовлетворяет известным условиям:

(i) $\boldsymbol{O} \subseteq \boldsymbol{Clos}^{\otimes}(\boldsymbol{O})$

(ii) $\boldsymbol{O_i} \subseteq \boldsymbol{O_j}$ влечет $\boldsymbol{Clos}^{\otimes}(\boldsymbol{O_i}) \subseteq \boldsymbol{Clos}^{\otimes}(\boldsymbol{O_j})$

(iii) $\boldsymbol{Clos}^{\otimes}(\boldsymbol{Clos}^{\otimes}(\boldsymbol{O})) = \boldsymbol{Clos}^{\otimes}(\boldsymbol{O})$

т.е. является оператором замыкания. С помощью оператора $\boldsymbol{Clos}^{\otimes}$ восстанавливается структура покрытия классами эквивалентности описаний прецедентов уже построенных классов сходства а также всего исходного множества $\mathbf{\Omega}$.

Таким образом демонстрируется, что задача извлечения зависимостей из имеющихся эмпирических данных может быть формализована как задача восстановления и последующего анализа взаимных пересечений классов эквивалентности $\boldsymbol{E}_{\boldsymbol{v}i}^{+}(\Omega,\otimes)$ и $\boldsymbol{E}_{\boldsymbol{v}j}^{-}(\Omega,\otimes)$, реконструируемых на структурных описаниях прецедентов противоположных знаков – примеров (Ω^{+}) и контрпримеров (Ω^{-}). (Система отношений эквивалентности, возникающая в рассматриваемой ситуации, оказывается производной как от выбранного формального уточнения содержательных представлений о сходстве в виде той или иной конкретной алгебраической операции $\otimes$, так и от текущей структуры имеющейся выборки прецедентов Ω).

Задача прогнозирования свойств новых прецедентов, в таком контексте формализуется как задача формирования множеств $\boldsymbol{Eq}^{+}(\Omega,\otimes)$ и $\boldsymbol{Eq}^{-}(\Omega,\otimes)$ всех классов эквивалентности $\boldsymbol{E}_{\boldsymbol{v}i}^{+}(\Omega,\otimes)$ и $\boldsymbol{E}_{\boldsymbol{v}j}^{-}(\Omega,\otimes)$ отдельно на каждом множестве прецедентов обоих знаков Ω^{+} и Ω^{-} и последующей проверки вложимости описания нового прецедента $\boldsymbol{O_0}$ хотя бы в один из классов эквивалентности $\boldsymbol{E}_{\boldsymbol{v}i}^{\alpha}(\Omega,\otimes)$ одного знака α ($\alpha\in\{+,-\}$) и, одновременно, невложимости ни в один из классов эквивалентности $\boldsymbol{E}_{\boldsymbol{v}j}^{-\alpha}(\Omega,\otimes)$ противоположного знака (т.е. невложимости для всех $\boldsymbol{E}_{\boldsymbol{v}j}^{-\alpha}(\Omega,\otimes) \in \boldsymbol{Eq}^{-\alpha}(\Omega,\otimes)$). При этом вложимость описания нового объекта $\boldsymbol{O_0}$ в класс эквивалентности $\boldsymbol{E}_{\boldsymbol{v}i}^{\alpha}(\Omega,\otimes)$, образованный прецедентами $\boldsymbol{O_{i1}},\boldsymbol{O_{i2}}, \ldots, \boldsymbol{O_{is}}$, понимается следующим образом:

$$\boldsymbol{O_{i1}} \otimes \boldsymbol{O_{i2}} \otimes \ldots \otimes \boldsymbol{O_{is}} = \boldsymbol{v}_i = (\boldsymbol{O_{i1}} \otimes \boldsymbol{O_{i2}} \otimes \ldots \otimes \boldsymbol{O_{is}}) \otimes \boldsymbol{O_0}$$

Определяя естественный порядок $\subseteq$ (по вложимости подмножеств исходно заданного множества прецедентов Ω^{α}) на множествах $\boldsymbol{Eq}^{\alpha}(\Omega,\otimes)$ восстанавливаемых из исходной выборки классов эквивалентности $\boldsymbol{E}_{\boldsymbol{v}i}^{\alpha}(\Omega,\otimes)$ каждого знака α ($\alpha\in\{+,-\}$), можно сформировать диаграммы частичного порядка:

$$\boldsymbol{D}^{\alpha} = < \boldsymbol{Eq}^{\alpha}(\Omega,\otimes), \subseteq >.$$

Структура и взаимосвязи изучаемых диаграмм противоположного знака $< \boldsymbol{Eq}^{\alpha}(\Omega,\otimes), \subseteq >$ и $< \boldsymbol{Eq}^{-\alpha}(\Omega,\otimes), \subseteq >$ определяют все основные особенности

алгоритмики порождения эмпирических зависимостей в рассматриваемом случае (необходимо выявлять и исключать из процесса прогноза свойств такие результаты вычисления сходства ⊗, когда для одного из классов эквивалентности сходство $\boldsymbol{\nu}_i$ примеров может совпадать со сходством $\boldsymbol{\nu}_j$ контрпримеров какого-либо класса эквивалентности прецедентов противоположного знака).

Каждая из диаграмм вида $< \boldsymbol{Eq}^{\alpha}(\Omega,\otimes), \subseteq >$ простыми действиями может быть дополнена до *решетки*: достаточно расширить множество $\boldsymbol{Eq}^{\alpha}(\Omega,\otimes)$ следующим набором новых элементов – самим множеством Ω^{α}, всеми его одноэлементными подмножествами и пустым множеством $\llcorner\varnothing\lrcorner$ прецедентов из Ω^{α}. (Теоретико-множественный случай чуть позднее - в **Разделе 4.2 Главы 4** - детально представлен ***Утверждениями 4.2.36 – 37***). Там же показано (***Примеры 4.2.38-39***), что порождаемые рассматриваемым способом решетки в общем случае (следуя, например, *Г.Гретцеру*) не являются дистрибутивными, т.к. могут содержать в виде подрешеток так называемые *пентагон* и *диамант* - известные решетки $\mathfrak{N}_5$ и $\mathfrak{M}_3$.

В **Разделе 2.2** приведено неформальное описание семантики так называемых правил индуктивного вывода Д.С.Милля (эвристики Ф.Бэкона-Д.С.Милля) - одного из базовых элементов задействованного в обсуждаемом подходе комплекса эвристик, формализуемых соответствующими логико-математическими средствами. Вместе с рядом других эвристик (в первую очередь - рассуждениями по аналогии в духе Д.Пойа и абдуктивными выводами в стиле Ч.С.Пирса) эта эвристика положена в основу ДСМ-метода автоматического порождения зависимостей из эмпирических данных.

В **Разделе 2.3** представлено общее описание и наивная семантика ДСМ-метода. Приведены ссылки на построенную В.К.Финном, Д.П.Скворцовым и О.М.Аншаковым дедуктивную имитацию процедурной конструкции теоретико-множественной версии ДСМ-метода средствами специального класса многозначных логик. Такая имитация позволяет показать (см. обсуждение *достаточности оснований* для принятия результатов обучения на прецедентах

в **Разделе 2.1**), что для каждого утверждения, которое может быть получено применением используемых в рамках ДСМ-метода правил правдоподобного вывода к конкретным исходным данным, в соответствующей (имитирующей ДСМ-рассуждения) дедуктивной теории средствами достоверного вывода может быть также получено аналогичное утверждение, и наоборот. (Таким способом демонстрируется корректность задействованного в рамках ДСМ-метода способа формализации используемых эвристик).

Глава 3 описывает процедурную структуру ДСМ-метода - формальное уточнение введенных в **Главе 2** базовых элементов развиваемого подхода:

- алгебраической формализации сходства как бинарной операции при работе с различными типами описаний прецедентов из исходной выборки **Ω**,
- детализации функциональных особенностей формализуемого комплекса эвристик (эмпирической индукции в стиле Ф.Бэкона-Д.С.Милля, выводов по аналогии, порождения абдуктивных объяснений в стиле Ч.С.Пирса и др.),
- обсуждению возможностей так называемых квази-аксиоматических теорий (специальных расширений аксиоматических теорий, использующих не только правила достоверного вывода, но также и правила правдоподобного вывода, назначение которых - порождение эмпирических зависимостей из фактов) как средства представления знаний в компьютерных системах ИАД, ориентированных на решение задач эмпирической индукции,
- детальному описанию используемых в ДСМ-ИАД инструментов анализа данных (так называемых решающих предикатов, правил правдоподобного вывода и стратегий формализованных ДСМ-рассуждений).

В **Разделе 3.1** рассмотрены варианты математической формализации понятия сходства на различных типах данных (множествах, графах, цепочках символов конечного алфавита, векторах числовых значений параметров, …), цель которых - порождение (в явной или неявной форме) соответствующего отношения *толерантности* (сходства).

Сперва дан обзор наиболее распространенных вариантов уточнения понятия сходства на базе математической теории меры, метрического анализа

близости (в том числе – и для операций с булевскими векторами) и, наконец, представлений о сходстве, формализуемом как алгебраическая операция. Затем рассмотрены варианты определения операции сходства для работы с множествами (в том числе - мультимножествами), продемонстрирована корректность использования операции $\cap$ пересечения множеств в качестве уточнения операции сходства (***Утверждения 3.1.2.3-5*** о том, что алгебры, которые формируются с использованием операции $\cap$ для работы с названными типами данных, являются *нижними полурешетками*). Операция сходства на графах понимается как операция выделения множества максимальных общих подграфов для пары (одноэлементных) множеств графов. Сходство на цепочках символов конечного алфавита формируется по той же схеме, что и сходство на графах.

Формализация сходства на векторах числовых параметров, дополняющих структурное описание прецедентов, выполняется по следующей схеме: представляется естественным предположить, что каждому структурному «носителю» изучаемых свойств анализируемых прецедентов соответствует область постоянства «физического механизма» обусловленности этих свойств в том числе и на принимающих числовые значения релевантных параметрах. Т.е. в множестве всех таких параметров существуют определенные подмножества релевантных целевому эффекту параметров, дополнительно характеризуемые релевантными эффекту числовыми значениями именно этих (выделенных) параметров. Односвязность области постоянства «физического механизма» причинности означает, что для любой пары C_1 и C_2 векторов числовых значений параметров, попавших в такую область, должна существовать возможность найти расположенный «между ними» вектор C_{12}, также попадающий в искомую область постоянства «механизма причинности». В свою очередь, понятие «между» может быть уточнено в терминах частичного порядка (числовых значений каждого их релевантных эффекту параметров). Наконец, понятие «между» одновременно для множеств релевантных параметров может быть уточнено сопутствующими изменениями всех их числовых значений вдоль так

называемой «оси монотонности» - упорядочения числовых значений какого-либо одного из релевантных параметров. Таким образом:

- упорядочив по возрастанию (или по убыванию) числовых значений одного из параметров R_i («оси монотонности») множество $\boldsymbol{C}^{\alpha}$ столбцов матрицы $\boldsymbol{X}^{\alpha}$ исходно заданных прецедентов, можно выделить все подмножества $\boldsymbol{C}_{\mu}$ множества столбцов-прецедентов $\boldsymbol{C}^{\alpha}$, на которых числовые значения, соответствующие каждому подмножеству $\boldsymbol{R}_{\mu}$ множества параметров $\boldsymbol{R}$, изменяются сопутствующим образом (ко-монотонно). В такой ситуации естественно считать *сходными* все прецеденты (столбцы), попадающие в конкретный участок ко-монотонности изменения числовых значений релевантных параметров (определяемый соответствующей монотонной матрицей $\mu=\boldsymbol{R}_{\mu}\text{x}\boldsymbol{C}_{\mu}$). При этом одновременно берутся максимальные по вложению множества строк $\boldsymbol{R}_{\mu}$, обеспечивающих ко-монотонные изменения числовых значений в строках матрицы $\mu=\boldsymbol{R}_{\mu}\text{x}\boldsymbol{C}_{\mu}$ на всем множестве образующих ее столбцов $\boldsymbol{C}_{\mu}$. Фактически

- *сходство* пары прецедентов характеризуется здесь некоторым множеством столбцов-прецедентов, располагающимся *между* этими (проверяемыми на сходство) столбцами в смысле упорядочения значений одного из параметров (строк исходной матрицы-универсума $\boldsymbol{X}^{\alpha}$) и выделения содержащей по крайней мере две строки м-матрицы $\mu=\boldsymbol{R}_{\mu}\text{x}\boldsymbol{C}_{\mu}$ в $\boldsymbol{X}^{\alpha}$, вдоль строк которой числовые значения релевантных параметров между сходными столбцами изменяются сопутствующим образом (ко-монотонно). *Классы эквивалентности* исходных прецедентов будут характеризоваться здесь максимальными по вложению (в универсуме $\boldsymbol{X}^{\alpha}$) множествами столбцов каждой такой м-матрицы.

Чтобы сформулировать корректное определение операции сходства на векторах числовых значений релевантных параметров

- сперва на м-матрицах задается операция склейки, результатом которой для двух м-матриц является такая третья (разумеется, если таковая существует), множество столбцов которой представляет объединение множеств столбцов

склеиваемых матриц, а множество строк есть (возможно пустое[5]) множество тех (общих для них) строк, вдоль которых на всем объединенном множестве столбцов числовые значения параметров изменяются ко-монотонно;

- затем на множествах м-матриц (порожденных в исходном универсуме $\boldsymbol{X}^{\alpha}$) строится бинарная операция $\pitchfork$ покомпонентной склейки[6]. Имеет место:

Лемма 3.1.2.11

На множестве $\boldsymbol{\mu}(\boldsymbol{X}^{\alpha})$ всех м-матриц исходно заданного универсума $\boldsymbol{X}^{\alpha}$ операция $\pitchfork$ удовлетворяет условиям коммутативности и ассоциативности.

□

Однако, приведен ***Пример 3.1.2.17***, демонстрирующий, что на произвольных подмножествах из $\boldsymbol{\mu}(\boldsymbol{X}^{\alpha})$ операция $\pitchfork$ не является идемпотентной. Тем не менее, если, стартовав с *одно-столбцовых подматриц*[7] матрицы-универсума $\boldsymbol{X}^{\alpha}$, (индуктивно) формировать специальное подмножество $\boldsymbol{Dom}(\boldsymbol{X}^{\alpha}) \subset 2^{\boldsymbol{\mu}(X)}$, порождая новые элементы $\boldsymbol{Dom}(\boldsymbol{X}^{\alpha})$ применением операции $\pitchfork$ только к уже имеющимся в $\boldsymbol{Dom}(\boldsymbol{X}^{\alpha})$ элементам, то для $\boldsymbol{Dom}(\boldsymbol{X}^{\alpha})$ операция $\pitchfork$ удовлетворяет и условию идемпотентности. Т.е. имеет место (определяющая корректность уточнения сходства на множестве $\boldsymbol{Dom}(\boldsymbol{X}^{\alpha})$ посредством операции $\pitchfork$):

Теорема 3.1.2.15

Алгебра $\wp = <\boldsymbol{Dom}(\mathbf{X}), \pitchfork>$ есть нижняя полурешетка.

□

В **Разделе 3.2.1** рассмотрена система логических языков для описания правил правдоподобного вывода, характеризующих процедурный механизм операций порождения и выявления взаимных зависимостей на множествах классов эквивалентности прецедентов (см. выше **Раздел 2.1**). Семантические особенности комплекса эвристик (индуктивного обучения, выводов по аналогии и абдуктивного объяснения), используемого в рамках формируемой формализации эмпирической индукции, обсуждаются в **Разделе 3.2.2**. Общие свойства правил правдоподобного вывода ДСМ-метода (условия их примени-

[5] В таком случае результат операции – пустое множество м-матриц.

[6] Каждая м-матрица из первого множества с каждой м-матрицей из второго.

[7] Т.е., фактически, с описаний исходных объектов-прецедентов.

мости, рациональность, инвариантность процедурного ядра при работе с различными типами данных и др.) анализируются в **Разделе 3.2.3**. Явный вид решающих предикатов и собственно правил правдоподобного вывода (в том числе – для операций с теоретико-множественными и числовыми данными) подробно представлен к **Разделах 3.3.1-3**. Показано, как предложенная в **Разделе 2.1** базовая процедурная схема формализации задачи эмпирической индукции *алгебраическими* средствами может быть описана однозначно интерпретируемыми *логико-математическими* средствами представленной в **Разделе 3.2.1** *системы логических языков*. Таким образом обеспечены возможности для обоснования корректности предлагаемой алгебраической формализации рассматриваемой версии задачи обучения на прецедентах в том числе и путем построения (средствами соответствующих формальных теорий см. выше – **Раздел 2.3**) проблемно-ориентированной *дедуктивной имитации*.

Центральная роль в работе отводится рассмотренным в **Главах 4** и **5** математическим результатам об оценках вычислительной сложности соответствующих переборных задач а также оптимизации алгоритмики предложенной процедурной конструкции обучения на прецедентах.

В **Главе 4** анализ комбинаторных свойств диаграмм вложимости классов эквивалентности начинается с изучения теоретико-множественного случая описания прецедентов. Исходные данные: алфавит $\mathbf{U}=\{a_1,a_2, \dots ,a_n\}$- универсум образующих, а $\mathbf{\Omega} = \{O_1, O_2, \dots , O_m\} \subseteq 2^{\mathbf{U}}\backslash\varnothing$ - множество объектов - слов (т.е. непустых множеств образующих), построенных над универсумом **U**. Операция сходства $\otimes$ пары слов O_i и O_j определяется как их теоретико-множественное пересечение $\cap$ (выдает общее для этих слов подмножество образующих алфавита **U**). На подмножествах алфавита **U** и множества всех слов $\mathbf{\Omega}$ определяются два отображения:

$$\boldsymbol{s} : 2^{\mathbf{\Omega}} \to 2^{\mathbf{U}} \quad \text{и} \quad \boldsymbol{g} : 2^{\mathbf{U}} \to 2^{\mathbf{\Omega}} \quad ,$$

такие, что: $\boldsymbol{s}$ каждому подмножеству слов из $\mathbf{\Omega}$ сопоставляет общее для них множество букв из **U**, а $\boldsymbol{g}$ каждому подмножеству букв из **U** сопоставляет под-

множество слов из **Ω**, в которые все эти буквы входят одновременно. Показано, что ***s*** и ***g*** на **U** и **Ω** представляют собою соответствия Галуа, а их произведения (***s***x***g***) и (***g***x***s***) порождают соответствующие замыкания Галуа. (Таким образом оператор замыкания ***Clos***$^{\otimes}$(***O***), введенный ранее в **Разделе 2.1**, уточняется здесь до оператора ***g***(***s***(***O***)) – замыкания Галуа на множестве слов **Ω**).

Показано, что для каждой задачи ***КлЭкв***(**U**,**Ω**,⊗) о порождении всех классов эквивалентности прецедентов из **Ω** может быть сформулирована двойственная ей задача ***КлЭкв***(**U***,**Ω***,⊗), где каждой букве (образующей) из **U** сопоставляется соответствующее слово из **Ω***, в свою очередь перечисляющее все слова из исходного множества прецедентов **Ω**, в которые входит эта буква, а двойственный алфавит **U*** кодирует своими образующими все слова из **Ω**. Взаимосвязи прямой и двойственной задач о порождении рассматриваемых нами классов эквивалентности демонстрирует

Теорема 4.2.5

В множестве ***Eq***(**U**,**Ω**,⊗) классов эквивалентности для задачи ***КлЭкв***(**U**,**Ω**,⊗) элемент, содержащий ровно k штук прецедентов из **Ω**, существует тогда и только тогда, когда в двойственном множестве классов ***Eq***(**U***,**Ω***,⊗) для задачи ***КлЭкв***(**U***,**Ω***,⊗) существует такой элемент, который сформирован ровно k образующими из двойственного алфавита **U***.

□

В соответствии со ***Следствием 4.2.6*** и ***Леммой 4.2.7*** аналогичное утверждение верно и при соотнесении классов ***Eq***(**U***,**Ω***,⊗) и ***Eq***(**U**,**Ω**,⊗). Описанный эффект имеет прямые аналогии с эффектами двойственности в линейном программировании и позволяет упростить доказательства ряда утверждений о вычислительной сложности переборных задач, связанных с формированием рассматриваемых диаграмм вложимости классов эквивалентности.

Далее в **Главе 4** сформулирован и доказан комплекс утверждений, которые определяют вычислительную сложность основных комбинаторных задач, характерных для порождения изучаемых диаграмм вложимости. В первую очередь – это задачи об оценках емкости (числа вершин) таких диаграмм, тру-

доемкости порождения их границ (максимальных и минимальных по вложению классов эквивалентности), сложности поиска в них классов эквивалентности с заданными структурными характеристиками (содержащих не менее\не более заданного числа исходных прецедентов; сформированных сходством, в котором не менее\ не более чем заданное число образующих исходного алфавита; сформированных ровно заданным числом прецедентов или же ровно заданным числом образующих и т.п.). Так сводимостью известной задачи о числе монотонных булевских функций, представленных в виде 2-КНФ, показано:

Теорема 4.2.10

Задача о числе классов эквивалентности, порождаемых на исходных множествах **U** и **Ω**, принадлежит классу #*PC* – т.е. перечислительно полна.

Таким образом, число порождаемых классов эквивалентности растет экспоненциально быстро с увеличением характерных размеров исходных данных. Далее показано (***Теоремы 4.2.12-13*** и ***Следствия 4.2.14-19***), что обе границы диаграмм $\boldsymbol{D}^{\alpha}$ рассматриваемого здесь типа порождаются процедурами полиномиальной вычислительной сложности.

Сводимостью задачи о выполнимости произвольной булевской функции показано (в т.ч. - доказательством ряда промежуточных утверждений - ***Лемм 4.2.21-26*** и ***Следствия 4.2.27***), что имеют место следующие два утверждения:

Теорема 4.2.20

Задача о наличии класса эквивалентности, который порождается ровно заданным числом образующих исходного алфавита **U,** принадлежит классу *NPC* – *NP*-полных переборных задач.

Следствие 4.2.28

Задача о классе эквивалентности, определяемая условием ***Теоремы 4.2.20***, принадлежит также и классу #*PC* перечислительно полных задач.

Таким образом показано, число даже таких специальных классов эквивалентности, вообще говоря, может расти экспоненциально быстро с ростом характерных размеров множеств **U** и **Ω**). Далее (с учетом ***Теоремы 4.2.5*** о двойственности) аналогичные результаты (о принадлежности тем же классам ком-

бинаторно трудных задач) доказаны и для задачи о поиске класса эквивалентности, который сформирован ровно заданным числом прецедентов из исходного множества $\boldsymbol{\Omega}$:

Следствие 4.2.29

Задача о наличии класса эквивалентности, который сформирован ровно заданным числом прецедентов из исходного множества $\boldsymbol{\Omega}$, принадлежит классам *NPC* (*NP*-полных) и *#PC* перечислительно полных переборных задач.

При переходе к другим (рассмотренным в **Разделе 3.1**) типам данных наследуется большинство полученных для теоретико-множественного случая экспоненциально быстро растущих оценок сложности вычислений. Так для *мультимножеств* (как показывает ***Следствие 4.3.5***) сохраняются результаты, установленные ***Теоремами 4.2.5, 4.2.10, 4.2.12-13, 4.2.20*** и ***4.2.35*** (включая соответствующие ***Леммы*** и ***Следствия***).

Далее доказана принадлежность классам *NPC* и *#PC* аналогичных задач о классах эквивалентности, содержащих заданное число прецедентов, и о числе классов эквивалентности при описании исходных прецедентов как *цепочками символов* конечного алфавита (***Следствие 4.3.6***), так и *графами* (***Следствие 4.3.7***).

Для представления исходных данных *векторами числовых значений* параметров показано, что задача о поиске какого-либо класса эквивалентности исходно заданных прецедентов полиномиально разрешима (***Теорема 4.3.10***), однако задача о числе таких классов (как и в ранее рассмотренных случаях) находится в классе *#PC* – перечислительно полных задач (***Теорема 4.3.11***). Тем не менее, в случае отсутствия совпадающих элементов в строках исходной матрицы-универсума $\boldsymbol{X}^{\alpha}$ с ростом ее размеров число классов эквивалентности растет полиномиально быстро (***Теорема 4.3.12***). Также, как и ранее, задачи о поиске границ множества классов эквивалентности и существовании классов эквивалентности, образованных не более\не менее чем заданным числом прецедентов, оказываются полиномиально разрешимыми (***Теоремы 4.2.14, 16, 18*** и ***Следствия 4.3.15,17***). Наконец, как и в случаях обработки ранее уже рассмотренных типов данных, доказана

Теорема 4.3.19

Задача о существовании класса эквивалентности, который образован ровно заданным количеством описываемых числовыми векторами исходных прецедентов, принадлежит классу *NP*-полных переборных задач – *NPC*.

Очевидный неформальный итог предпринятого в **Главе 4** анализа: полученные (причем – однотипные для различных вариантов описания анализируемых примеров и контрпримеров) *оценки сложности* комбинаторных задач, возникающих при реконструкции классов эквивалентности исходно заданных прецедентов, - основной аргумент в пользу *актуальности* разработки проблемно-ориентированных *методов оптимизации перебора* вариантов, характерного для развиваемых средств формализации эмпирической индукции.

В **Главе 5** обсуждается разработанный автором диссертационного исследования метод дискретной оптимизации перебора вариантов при восстановлении (по исходно заданной выборке) диаграмм вложимости классов эквивалентности прецедентов, описываемых множествами признаков (теоретико-множественный случай). Вводится понятия базиса замыканий Галуа – замыканий (вида $\boldsymbol{s}(\boldsymbol{g}(a))$ - см. ранее, **Раздел 4.2**) всех однобуквенных подмножеств алфавита **U**. Предложена следующая рекурсивная процедура:

даны: - множество примеров $\boldsymbol{\Omega}$, построенных на алфавите $\boldsymbol{U}=\{a_1,a_2,\ldots,a_n\}$ и
- непустое подмножество $\mathbf{u}=\{u_1,u_2,u_3,\ldots,u_k\}$ алфавита $\boldsymbol{U}$.

(**i**) По $\boldsymbol{\Omega}$ и $\boldsymbol{U}$ строится множество замыканий всех одноэлементных подмножеств для $\boldsymbol{U}$:

$$\boldsymbol{B\_GC}_{s*g}(\mathbf{U},\boldsymbol{\Omega}) = \{[\{a_1\}],[\{a_2\}],\ldots,[\{a_n\}]\}.$$

(**ii**) В **u** выбирается u_1 – порождающее максимальный элемент в $\boldsymbol{B\_GC}_{s*g}(\mathbf{U},\boldsymbol{\Omega})$.

(**iii**) В $\boldsymbol{\Omega}$ выделяется подмножество $\boldsymbol{\Omega}(u_1)$ всех содержащих $[u_1]_{U,\boldsymbol{\Omega}}$ примеров.

(**iv**) Из каждого входящего в $\boldsymbol{\Omega}(u_1)$ примера удаляются образующие из $[u_1]_{U,\boldsymbol{\Omega}}$ и формируется новое – модифицированное - множество примеров $\boldsymbol{\Omega}^*(u_1)$. В результате порождаются: усеченное множество примеров $\boldsymbol{\Omega}^*(u_1)$ и использованный при его формировании усеченный алфавит $\boldsymbol{U}(u_1)$.

(**v**) Из текущего **u** удаляются все образующие, вошедшие в $[u_1]_{U,\boldsymbol{\Omega}}$, после чего

(**vi**) возвращаемся к построению $\boldsymbol{B\_GC}_{s*g}$, но теперь уже на усеченных множествах $\boldsymbol{\Omega}^*(u')$ и $\boldsymbol{U}(u')$, и повторяем процедуру (**i**)-(**v**) на усеченных множествах $\boldsymbol{\Omega}^*(u')$ и $\boldsymbol{U}(u')$ рекурсивно до момента, когда в текущем состоянии

множества **u** больше не останется ещё не удалённых образующих. Последовательность $u_1,u_2, \ldots, u_l$ (где $l \le k$) элементов из множества **u** образующих исходного алфавита ***U***, удаляемых на каждом из шагов предпринимаемой рекурсии, - T(**u**,**Ω**,***U***) - назовем *траекторией усечения* исходного **u**.

Полезные свойства описанной рекурсивной процедуры представляет:

Теорема 5.2.3

Множество образующих **u** замкнуто относительно множества примеров **Ω** в алфавите ***U*** тогда и только тогда, когда объединение всех замыканий $[u']_{U,\Omega}$ вдоль траектории T(**u**,**Ω**,***U***) совпадает с **u**:

$$[\mathbf{u}] = \mathbf{u} \quad \textit{тогда и только тогда, когда} \quad \Big(\bigcup_{u'_i \in T(u,\Omega,U)} [u'_i] \Big) = \mathbf{u}.$$

□

Показано, что каждая диаграмма вложимости рассматриваемого типа может быть представлена объединением поддиаграмм специального вида - так называемых псевдо-деревьев (в корне каждого из которых находится описание одного из исходно заданных прецедентов, а остальные вершины содержат лишь входящие в корень элементы – образующие исходного алфавита ***U***).

В рассматриваемых диаграммах выделены архитектурные фрагменты – линейные ветви частичного порядка (фрагменты типа α), комбинаторно сложные фрагменты булевских гиперкубов (фрагменты типа β) и их комбинации – фрагмент типа α, наложенный на фрагмент типа β (фрагмент типа γ). Предложена технология порождения приближенных описаний псевдо-деревьев, стартующая с порождения так называемого *каркаса* – диаграммы взаимной вложимости всех элементов базиса Галуа, соответствующих всем элементам множества образующих описания прецедента (из исходного множества **Ω**), лежащего в корне текущего псевдо-дерева. Установлены следующие свойства каркасов:

Утверждение 5.2.11

Каркас любого псевдо-дерева (как и каркас всей восстанавливаемой по исходным данным диаграммы) порождается полиномиально сложной процедурой.

Утверждение 5.2.12

Каркас однозначным образом сопоставлен соответствующему псевдо-дереву.

□

Показано (***Утверждения 5.2.7*** и ***5.2.13***), что в процессе построения каркаса имеется возможность быстро (т.е. полиномиально сложными вычислениями) не только выделять в соответствующем псевдо-дереве линейные цепочки частичного порядка (архитектурные фрагменты типа α), но и диагностировать наличие в нем экспоненциально сложных архитектурных фрагментов типа β.

Найдены (***Утверждение 5.2.15***) полиномиально быстро проверяемые необходимые и достаточные условия, при выполнении которых обнаруженный при формировании каркаса этого псевдо-дерева комбинаторно-сложный фрагмент типа β представляет собою полный гиперкуб (на некотором подмножестве образующих исходного алфавита ***U***).

Предложен алгоритм **НПС**[8], обеспечивающий целенаправленный исчерпывающий перебор всех элементов соответствующих диаграмм вложимости классов эквивалентности. Корректность алгоритма **НПС** демонстрирует

Утверждение 5.3.1

Алгоритм НПС обеспечивает полноту и точностью при восстановлении диаграммы вложимости сходств, порождаемых на исходных множествах ***U*** и **Ω** :

1) порождает все принадлежащие ей сходства (с указанием для каждого из них всех ближайших к нему по вложимости элементов восстанавливаемой диаграммы), и
2) не порождает никаких других (не являющихся элементами восстанавливаемого множества сходств) объектов.

□

Для случая, когда соответствующий комбинаторно-сложный фрагмент псевдо-дерева (фрагмент типа β) составляет лишь некоторую (требующую восстановления) часть такого гиперкуба, разработан *метод последовательных приближений* (целенаправленного порождения все более сложных описаний архитектурных фрагментов подобного типа), базовые компоненты которого:

- проблемно-ориентированная «сжимающая» процедура на множестве исходно заданных прецедентов, позволяющая последовательно переходить к

[8] Алгоритм **Н**аправленного **П**еребора **С**ходств (порождающих восстанавливаемые классы эквивалентности).

целенаправленно выбираемым подзадачам (реконструкции с *заданной степенью точности* приближенного описания целенаправленно выбираемых архитектурных фрагментов типа β) при восстановлении каждого анализируемого псевдо-дерева, и

- специальная (ограниченная сверху по формируемым ею результатам точным описанием целевого псевдо-дерева и всей диаграммы вложимости) процедура, монотонно расширяющая текущее приближенное описание реконструируемого псевдо-дерева вновь порождаемыми при анализе подзадач (т.е. при реконструкции соответствующих гиперкубов - см. «сжимающую» процедуру выше) архитектурными фрагментами.

Предложенный метод позволяет порождать в первую очередь те фрагменты восстанавливаемых диаграмм, которые полезны для прогнозирования свойств новых прецедентов. Далее (разумеется, - за счет все более и более объемных вычислений) может быть порождено и точное описание каждой такой диаграммы. Показано, что этот метод позволяет организовать процесс порождения целевых диаграмм вложимости в режиме *параллельных* вычислений.

В **Разделе 5.4** представлено строгое обоснование корректности разработанного метода последовательных приближений (так называемого *приближенного* ДСМ-метода). Описан механизм управления порождением (в том числе – в режиме *параллельных* вычислений) последовательных приближений заданной точности для восстанавливаемых диаграмм. Вводится понятие *каркаса* $GK^{(i)}_{\boldsymbol{D}}$ *ранга* i - диаграммы взаимной вложимости элементов множества замыканий $\boldsymbol{s^*g}$ всех i-элементных подмножеств алфавита $\boldsymbol{U}$

$$\boldsymbol{B}^{(i)}\_\boldsymbol{GC}_{s^*g}(\mathbf{U},\boldsymbol{\Omega}) = \{\boldsymbol{s^*g}\ (\{a_{j_1}, a_{j_2}, \dots, a_{j_i}\}) \mid j_1, j_2, \dots, j_i \in \{1,2, \dots ,n\}\}$$

Механизма *управления точностью* последовательных приближений основан на утверждении о представимости восстанавливаемых диаграмм вложимости объединениями каркасов всех допустимых рангов:

Теорема 5.4.2.

Для заданных алфавита **U** и множества примеров **Ω** диаграмма $\mathbf{D}\_\boldsymbol{GC}_{s^*g}(\mathbf{U},\boldsymbol{\Omega})$ представима объединенной диаграммой

$$\mathbf{D\_GC}_{s*g}(\mathbf{U},\Omega) = < (\bigcup_{i=1}^{n} \boldsymbol{B}^{(i)}\_\boldsymbol{GC}_{s*g}(\mathbf{U},\Omega)), \subseteq >,$$

сформированной из каркасов $GK^{(i)}_{\mathbf{D}}$ замыканий всех возможных рангов ***i*** на алфавите **U** (где |**U**| = n) и наследующей отношение непосредственной вложимости элементов в формирующих ее каркасах[9].

При описании каждой соответствующей диаграммы степень *точности приближения* регулируется выбором целевых[10] (восстанавливаемых в приоритетном порядке) архитектурных фрагментов типа β (т.е. фрагментов гиперкубов, идентифицированных при построении каркасов ранга 1) а также ограничениями по числу используемых при их реконструкции каркасов следующих рангов (выбором для каждого из таких фрагментов типа β своей необходимой точности описания). Важно, что все эти фрагменты можно восстанавливать в *параллельном* режиме вычислений.

Представленные в **Главе 5** результаты имеют центральное значение для всей рассматриваемой диссертационной работы, т.к. здесь сформулирована алгоритмически корректная и строго обоснованная процедурная конструкция, позволяющая (при наличии ряда «тяжелых» ограничений в части сложности вычислений – см. **Главу 4**) при порождении эмпирических зависимостей из данных и прогнозировании свойств новых объектов оперировать выборками прецедентов не ограничиваемого размера. (Для решения прикладной задачи эффективный размер выборки фактически определяется возможностями вычислительной установки, на которой будет организован интеллектуальный анализ соответствующих данных средствами *приближенного* ДСМ-метода, в свою очередь выполняемого в режиме *параллельных* вычислений).

---

[9] В том числе - с учетом используемого при формировании каркасов отношения непосредственной вложимости соответствующих множеств образующих из алфавита U, принадлежащих при этом каркасам разных рангов.

[10] Имеющих отношение к прогнозированию свойств новых прецедентов, контролю достаточности оснований для принятия результатов прогноза и т.п. .

В **Главе 6** рассматриваются некоторые дополнительные математические особенности предложенного инструментария ИАД. Представлены эффекты немонотонности правдоподобного вывода в квази-аксиоматических теориях, формализующих ДСМ-ИАД (***Пример 6.1.1***). Доказана *функциональность* отношения причинности, формальное уточнение которого построено предложенными логико-алгебраическими средствами (***Утверждение 6.2.2***). Предложен ряд процедурных соображений о способах контроля устойчивости эмпирических зависимостей, порождаемых развиваемым в диссертационном исследовании процедурным механизмом, при расширении текущей выборки прецедентов.

В **Приложениях** дан обзор применения интеллектуальных компьютерных систем, основанных на предложенном комплексе математических моделей методов и алгоритмов, при решении задач ИАД в ряде практически значимых прикладных предметных областей:

- анализе и прогнозировании как полезных (в том числе - лекарственных), так и антипродуктивных (токсичности, канцерогенности, …) свойств физиологически активных химических соединений (**П.2.1 -3**);
- автоматической обработке больших коллекций полнотекстовых документов (**П.3.1-4**);
- обработке данных геологических исследований и прогнозировании нефтегазоносности территорий (**П.4.1-2**);
- предконкурсной экспертизе сложных технических проектов (**П.6**)

и др. .

Основные математические результаты диссертационной работы представлены в **Разделах 2.1** и **3.1**, **Главах 4** и **5**. Обсуждению содержательных характеристик формализуемых эвристик (индуктивного обучения, выводов по аналогии и абдуктивного объяснения) посвящены **Разделы 2.2**, **3.2** и **6.4**.

Работоспособность развиваемого подхода при решении прикладных задач интеллектуального анализа данных продемонстрирована в **Приложениях**.

Основные результаты и выводы

Полученные в рассматриваемой диссертационной работе результаты позволяют при создании компьютерных систем анализа данных использовать современные методы искусственного интеллекта, в частности - корректные формализации исследовательских эвристик, формируемые средствами алгоритмических конструкций в том числе и высокой вычислительной сложности. Таким образом, имеются основания говорить о реализации идей, моделей и методов теоретической информатики в компьютерном инструментарии решения сложных задач анализа данных в прикладных предметных областях.

Получены следующие результаты:

1. Исследованы возможности организации обучения на прецедентах и автоматического извлечения зависимостей вида **ОБЪЕКТ=>СВОЙСТВА** из *данных нечислового характера*, базирующиеся на использовании комплекса эвристик *эмпирической индукции* (индуктивного обучения, выводов по аналогии и абдуктивного объяснения).
2. Разработана *алгебраическая формализация* для используемой процедурной конструкции извлечения зависимостей из структурных описаний прецедентов, исходно заданных для обучения.
3. Предложенная алгебраическая формализация обеспечивает *унифицированные* возможности оперировать *различными* типами структурных описаний исходных данных – множествами (в том числе - мультимножествами), графами, цепочками символов конечного алфавита и др., дополненными *числовыми значениями* существенных параметров.
4. Идентифицирован и изучен *специальный класс комбинаторных объектов* - диаграмм взаимной вложимости классов эквивалентности описаний объектов-прецедентов, - *определяющих особенности алгоритмики* порождения эмпирических зависимостей в рамках предлагаемой формализации эмпирической индукции.

5. Процедурная конструкция ДСМ-метода автоматического порождения зависимостей расширена на *новые типы данных* (структурные описания *нечисловых объектов*, дополненные *числовыми значениями* существенных параметров).
6. Получены *оценки сложности вычислений*, требуемых для решения основных переборных задач, которые возникают в процессе порождения диаграмм вложимости классов эквивалентности при обработке *различных типов данных* (вариантов описаний исходно заданных прецедентов). Доказана принадлежность соответствующих переборных задач ряду известных классов комбинаторно-трудных проблем.
7. Разработан метод *дискретной оптимизации* и *управления* комбинаторным *перебором* вариантов при формировании рассматриваемых диаграмм и анализе переносимости порождаемых эмпирических зависимостей на описания новых - ранее еще не изученных - прецедентов (*метод последовательных приближений* при формировании анализируемых диаграмм вложимости).
8. Предложен вариант такого метода последовательных приближений, реализуемый в режиме *параллельных вычислений*, что позволяет *снять ограничения* на *размеры* эффективно обрабатываемых *исходных коллекций прецедентов* а также существенно расширить область практических приложений развиваемой техники извлечения зависимостей из эмпирических данных.
9. Предложено строгое обоснование *корректности* процедурной конструкции разработанного метода последовательных приближений.
10. *Практическая значимость и эффективность* предложенных математических моделей, методов и алгоритмов подтверждена результатами, полученными при решении прикладных задач в различных предметных областях (при решении прикладных задач интеллектуального анализа данных средствами ДСМ-метода автоматического порождения зависимостей).

В разделе Перспективы дальнейших разработок представлены направления дальнейшего развития разработанных в диссертации математических мо-

делей и алгоритмических конструкций, которые позволяют сформировать теоретические основы новых технологий и расширить как процедурные возможности, так и область эффективного применения в практически важных приложениях осуществляемого представленными средствами интеллектуального анализа данных (развитие моделей и средств «опознания» устойчивых эмпирических зависимостей - *эмпирических закономерностей*; оперативный учет динамики изменений в описаниях исходно заданных прецедентов; развитие средств ранжирования зависимостей, порождаемых из исходной выборки прецедентов с учетом структуры решеток правил правдоподобного вывода ДСМ-метода, и др.).

Публикации по теме диссертации в изданиях из списка ВАК:

1. Забежайло М.И. Ненаследуемость эмпирического противоречия в ДСМ-методе и немонотонные рассуждения // Научно-техническая информация, Сер.2. - 1984. - N11. - С.14-17.
2. Забежайло М.И К проблеме расширения ДСМ-метода автоматического порождения гипотез на данные с числовыми параметрами // Научно-техническая информация. Сер.2. - 1992. - N11. - С.11-21
3. Забежайло М.И. К проблеме расширения ДСМ-метода автоматического порождения гипотез на данные с числовыми параметрами II // Научно-техническая информация. Сер.2. - 1993. - N2. - С.6-16.
4. Забежайло М.И. Формальные модели рассуждений в принятии решений: приложения ДСМ-метода в системах интеллектуального управления и автоматизации научных исследований // Научно-техническая информация, Сер.2. - 1996. - N5-6. - С.20-32.
5. Забежайло М.И. О комбинаторной природе одной задачи оптимизации // Научно-техническая информация, Сер.2. - 1996. - N1. - С.19-27.
6. Забежайло М.И. К проблеме формализации метода геологических аналогий. // Известия РАН. Сер. Теория и системы управления. - 1997. - N2.- С.151-164.

7. Забежайло М.И., Емеленко М.Н., Липчинский Е.А., Максин М.В. К пониманию термина “прикладная семиотика” // Научно-техническая информация, Сер.2 - 1998. - N1. – С. 11-18.
8. Забежайло М.И. К проблеме автоматического понимания полнотекстовых документов в информационном поиске. // Известия РАН. Сер. Теория и системы управления. - 1998. - N5. - С.167-176.
9. Забежайло М.И. Интеллектуальный анализ данных – новое направление развития информационных технологий // Научно-техническая информация, Сер. 2.- 1998. - №8. - С. 6-17.
10. Забежайло М.И. О функциональности отношения причинности, используемого в ДСМ-рассуждениях // Научно-техническая информация. Сер.2. - 2013. – N7.- С.1-7.
11. Забежайло М.И. О некоторых возможностях управления перебором в ДСМ-методе // Искусственный интеллект и принятие решений. – 2014. – Часть I: № 1, С.95 -110.
12. Забежайло М.И. О некоторых возможностях управления перебором в ДСМ-методе // Искусственный интеллект и принятие решений. – 2014. – Часть II: № 3, С.3 - 21.
13. Забежайло М.И., Синякова Е.В. К вопросу об «интеллектуальности» интеллектуального анализа данных // Научно-техническая информация. Сер. 2. - 2014. - № 3. - С. 1-9.
14. Забежайло М.И. Приближенный ДСМ-метода на примерах // Научно-техническая информация. Сер.2. – 2014. – №10, С. 1-12.
15. Забежайло М.И. К вопросу о достаточности оснований для принятия результатов интеллектуального анализа данных средствами ДСМ-метода // Научно-техническая информация. Сер.2. – 2015. – №1. – С.1-9.
16. Забежайло М.И. О некоторых оценках сложности вычислений в ДСМ-рассуждениях // Искусственный интеллект и принятие решений. - 2015. – Часть I: №1. – С. 3-17.
17. Забежайло М.И. О некоторых оценках сложности вычислений в ДСМ-рассуждениях // Искусственный интеллект и принятие решений. - 2015. – Часть II: №2 - С. 3-17.

Другие публикации по теме диссертации:

18. Zabezhailo M.I. et al. Reasoning Models for Decision Making: Applications of JSM-Method for Intelligent Control Systems//Architectures for Semiotic Modeling and Situation Analysis in Large Complex Systems. - Proc. of the Workshop of 10th (1995) IEEE Symp. on Intelligent Control (Eds.: J.Albus, A.Meystel, D.Pospelov, T.Reader). - 27 - 29 August 1995, Monterey, CA../ AdRem, Inc., 1995. - Pp. 99-108.
19. Zabezhailo M.I. To the Scalable Technology of Automated Document Understanding Based on Quasi-Axiomatic Theories.// Proc. 1997-th International Conference on Intelligent Systems and Semiotics “A Learning Perspective” - ISAS’97 (NIST, Gaithersburg, MD, September 22-25, 1997). - NIST Special Publications. - N918. - 1997. - Pp.100-102.
20. Забежайло М.И., Финн В.К., Козлова С.П., Катамадзе Т.Г., Авидон В.В., Рабинков А.А. Об одном методе автоматического формирования гипотез и его программной реализации // НТИ, Сер.2. - 1982. - N4. - С.20-26.
21. Забежайло М.И., Финн В.К., Авидон В.В., Катамадзе Т.Г., Блинова В.Г., Бодягин Д.А., Рабинков А.А. Об экспериментах с базой данных с неполной информацией посредством ДСМ-метода автоматического порождения гипотез // НТИ, Сер.2. - 1983. - N2. - С.28-32.
22. Забежайло М.И., Ивашко В.Г., Кузнецов С.О., Михеенкова М.А., Хазановский К.П., Аншаков О.М. Алгоритмические и программные средства ДСМ-метода автоматического порождения гипотез. // НТИ, Сер.2. - 1987. - N10. - С.1-13.
23. Забежайло М. И. Новые информационные технологии и системы НТИ // НТИ, Сер. 2. — 1990. — №5. — С. 2—9.
24. Забежайло М.И. Некоторые тенденции в развитии интеллектуальных систем // Программные продукты и системы. - 1990. - N 4. - С.86-94.
25. Забежайло М.И. Интеллектуальные системы и задача восстановления эмпирических зависимостей структурно-числового характера // Итоги науки и техники. Сер." Информатика". Т.15 "Интеллектуальные информационные системы" - М: ВИНИТИ. - 1991. - С. 102-114.
26. Забежайло М.И. Новые информационные технологии в научных исследованиях и технологических разработках // НТИ. Сер. 2.- 1992.- № 6.-С.1-11.
27. Забежайло М.И. Интеллектуальные системы: на пути к новым поколениям. // Новости искусственного интеллекта. N1, 1992. - С.8-24.
28. Забежайло М.И. (Zabezhailo M.I.) The extension of the JSM-method: plausible reasoning dealing with numeric data // Новости Искусственного Интеллекта. Специаль-

ный выпуск: Международная конференция по искусственному интеллекту "Восток-Запад - 93" (Artificial Intelligence News. Special issue: EWAIC-93, Moscow, 7-9 Sept. 1993). 1993.

29. Забежайло М.И. Формальные модели рассуждений в принятии решений: приложения ДСМ-метода в системах интеллектуального управления и автоматизации научных исследований // НТИ, Сер.2. - 1996. - N5-6. - С.20-32.
30. Забежайло М.И. О комбинаторной природе одной задачи оптимизации // НТИ, Сер.2. - 1996. - N1. - С.19-27.
31. Забежайло М.И., Емеленко М.Н., Липчинский Е.А., Максин М.В. К разработке платформенно-независимой версии программной системы, реализующей ДСМ-метод автоматического порождения гипотез// Труды 3-й Международной конференции “НТИ-’97: Информационные ресурсы, интеграция, технологии” (Москва, 26-28 ноября 1997 г.), М.: ВИНИТИ, Часть 1, С.91-93.
32. Забежайло М.И. Информационный поиск в полнотекстовых документах: некоторые проблемы и перспективы. // Новости искусственного интеллекта. - 1998. - N.1.- С. 24-59
33. Забежайло М.И. Data Mining & Knowledge Discovery in Data Bases: предметная область, задачи, методы и инструменты // Труды 6-ой национальной конференции по искусственному интеллекту с международным участием. - Пущино. - 1998. - Том 1. - С. 592-600.
34. Zabezhailo M.I., Finn V.K., Leybov A.E.,Melnikov N.I., Pankratova E.S. CBR-technology in the chemical safety control // Proc. EWCBR’94 (France, November 7-11, 1994).- 1994.
35. Zabezhailo M.I., Finn V.K. Intelligent Information Systems. - International Forum on Information and Documentation (FID, The Hague, Netherlands). - 1996. - V21. - N2. - Pp.21-31.
36. Zabezhailo M.I. On the Application of the Semiotic Modelling Technique in the Problem of Clearing Payments Control.// Proc. 12th European Conference on Artificial Intelligence (ECAI’96, Budapest, 11-16 August, 1996). - Workshop N30 “Applied Semiotics”. Pp.17-21.
37. Zabezhailo M.I., Finn V.K., Gergely T. JSM-method as an Instrument for Semiotic Modelling: Application to Geological Forecasting.// Proc. 12th European Conference on Artificial Intelligence (ECAI’96, Budapest, 11-16 August, 1996). - Workshop N30 “Applied Semiotics”. Pp.13-16.