

*На правах рукописи*



**Чебанов Дмитрий Константинович**

**ПРОГНОЗИРОВАНИЕ РЕМИССИИ ИЛИ РЕЦИДИВА  
У ОНКОЛОГИЧЕСКИХ БОЛЬНЫХ,  
РЕАЛИЗУЕМОЕ В ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЕ**

**05.13.17 – Теоретические основы информатики**

Автореферат диссертации на соискание ученой степени  
кандидата технических наук

Москва - 2021

Работа выполнена в Федеральном исследовательском центре «Информатика и управление» Российской академии наук (ФИЦ ИУ РАН).

**Научный руководитель:** **Забейайло Михаил Иванович**  
доктор физико-математических наук, Федераль-  
ный исследовательский центр «Информатика и  
управление» Российской академии наук, глав-  
ный научный сотрудник отдела №16

**Официальные оппоненты:** **Еремеев Александр Павлович,**  
доктор технических наук, профессор, Нацио-  
нальный исследовательский университет «Мос-  
ковский энергетический институт», заведующий  
кафедрой Прикладной математики

**Королева Мария Николаевна,**  
кандидат технических наук, Национальный ис-  
следовательский университет «Московский гос-  
ударственный технический университет им.  
Н.Э. Баумана», доцент кафедры "Компьютерные  
системы автоматизации производства"


**Ведущая организация:** Федеральное государственное бюджетное учре-  
ждение науки «Институт автоматизации и процес-  
сов управления» Дальневосточного отделения  
Российской академии наук

Защита состоится 23 декабря 2021 г. в 15:00 часов на заседании диссертацион-  
ного совета Д 002.073.05 на базе 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 15 ноября 2021 года.

Ученый секретарь  
диссертационного совета Д 002.073.05  
кандидат технических наук

 Рейер Иван Александрович

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Одним из основных направлений искусственного интеллекта (ИИ) как области исследований и разработок является интеллектуальный анализ данных (ИАД), который реализуется с помощью интеллектуальной системы (ИС), и представляет собой процесс поиска новых знаний - закономерностей в исходных данных. Наиболее оправданно применение ИАД в тех областях, где отсутствует строгая формализация данных, а также нет однозначного общепринятого алгоритма их обработки. К таким областям относятся медицина, биология, химия, социология, и другие научные дисциплины.

Проблематика формирования (конструктивного порождения) нового знания из уже имеющихся знаний и постоянно накапливаемых эмпирических данных (knowledge discovery) – одно из актуальных и практически значимых направлений исследований и разработок в области ИИ.

Пример актуальной задачи такого типа:

i) Анализируемая проблема:

- исследуется «механизм» возникновения определенного явления\эффекта;
- влияние на эффект оказывает выделенное большое множество факторов;
- в той или иной конкретной ситуации каждый из таких факторов в отдельности может быть либо активен («влияет»), либо не активен («не влияет»);
- все возможные комбинации активных и неактивных факторов распадаются на два класса – «безвредные» (когда изучаемый целевой эффект не наблюдается) и «вредоносные» (когда наличествует целевой эффект).

ii) Исходные данные:

- заданное множество «примеров» (некоторых «вредоносных» комбинаций факторов влияния) и
- заданное множество «контр-примеров» (некоторых «безвредных» комбинаций факторов влияния).

iii) Найти:

- все такие подмножества факторов влияния, которые содержат все те и только те факторы («замкнутые» - несокращаемые и непополняемые -

подмножества факторов), которые определяют «вредоносность» имеющихся примеров и безвредность имеющихся контрпримеров.

Пример актуальной прикладной задачи такого типа: выявление механизмов возникновения меланомы (рака кожи человека), определяемых мутациями отдельных генов (из множества в несколько сотен релевантных этой проблеме генов), в ситуации, когда некоторые подмножества индивидуальных мутаций – вредоносны (ведут к возникновению меланомы), а некоторые – нет.

Современная онкология является одной из наиболее наукоемких отраслей медицины. Онкологическое заболевание характеризуется сбоями в регуляции процессов жизнедеятельности клеток, и возникает в результате комбинации различных факторов: как внутренних, проявляющихся в организме, так и внешних. Процесс развития опухоли всегда происходит в результате нарушений в работе ДНК. Далеко не все функции генов, составляющих ДНК, в настоящий момент известны науке, соответственно, речь идет о плохо формализованной области знаний. В данном случае применение методов ИАД является оправданным.

Одним из инструментов ИАД, который был успешно применен к задачам предсказания возникновения и прогнозирования течения онкологического заболевания [1,2], является ДСМ-метод автоматизированной поддержки исследований (ДСМ-метод АПИ) [3,4,5,7], который реализует обнаружение закономерностей в сложноструктурированных эмпирических данных, в неявном виде содержащих причинно-следственные зависимости. Он обеспечивает формализацию знаний предметной области средствами многозначной логики, для чего обобщает в гипотезах информацию, полученную из обучающей выборки, затем применяет эти гипотезы для предсказания исследуемого эффекта у неизученных объектов, а также имеет критерий достаточного основания для принятия результатов правдоподобного вывода. ИС на основе ДСМ-метода АПИ реализует ИАД средствами правдоподобных рассуждений вида «индукция - аналогия - абдукция».

**Актуальность** диссертационного исследования обусловлена необходимостью решения прикладных и фундаментальных задач в онкологии при помощи ИАД. Так, прогнозирование заболевания важно в клинической практике для планирования лечебного процесса и превентивной терапии, а открытие ключевых механизмов возникновения и развития злокачественной опухоли необходимо для создания новых препаратов и методов лечения.

**Предметом исследования** диссертации являются методы ИАД, реализуемые при помощи ИС на основе ДСМ-метода АПИ и их практическая применимость для решения задач онкологии.

**Цель диссертационной работы:**

- разработка на базе ДСМ-метода автоматизированной поддержки научных исследований (АПИ) методов и средств автоматизированного порождения новых знаний из постоянно накапливаемых эмпирических данных и знаний (knowledge discovery);
- применение ДСМ-метода АПИ в качестве процедурной платформы для автоматизированной компьютерной поддержки исследований в онкологии;
- разработка реализующей ДСМ-метод АПИ интеллектуальной компьютерной системы;
- экспериментальная апробация разработанной компьютерной интеллектуальной системы (обоснование ее практической применимости к области анализа онкологических данных).

**Задачи исследования, определяемые целями диссертационной работы:**

- идентифицировать требующие применения интеллектуального анализа данных (ИАД) актуальные задачи исследуемой предметной области – анализа постоянно накапливаемых разнородных онкологических данных с целью поддержки принятия врачебных решений;
- разработать результативные средства представления данных, составляющих базу фактов (БФ), в том числе генетических данных пациентов с онкологическим диагнозом;

- реализовать в разрабатываемом программном инструментарии ИАД основные стратегии ДСМ-метода АПИ (простое сходство, сходство с запретом на контр-примеры, сходство-различие, сходство-различие с запретом на контр-примеры), а также их объединение;
- реализовать необходимые вспомогательные процедуры ДСМ-решателя: выявление эмпирических закономерностей – таких эмпирических зависимостей (формируемых в процессе компьютерных рассуждений гипотез), которые сохраняются при всех перестановках последовательных расширений БФ и обеспечивают правильные предсказания на тестовых данных, проверку непротиворечивости наборов гипотез, фильтры гипотез для удобства их перебора и проверки экспертом и др.;
- разработать реализующую ДСМ-рассуждение интеллектуальную систему для анализа онкологических данных, имеющую комфортный (в том числе – для пользователей, не являющихся ИТ-экспертами) графический пользовательский интерфейс;
- реализовать функции направленной навигации в диаграмме гипотез в рамках концепции приближенного ДСМ-метода АПИ, позволяющие пользователю в интерактивном режиме быстро (за короткое время) проверять текущую БФ на предмет наличия в ней (пусть и в неявном виде) содержательно интерпретируемых гипотез;
- провести апробацию разработанной ИС посредством проведения экспериментов на реальных исходных данных онкологических пациентов, содержащих в том числе и их генетические, и иммунные персоналифицированные характеристики.

Для решения вышеперечисленных задач использовались следующие **методы**: неклассическая (многозначная) логика, правдоподобные рассуждения, порождающие гипотезы о причинах эффектов и гипотезы о предсказаниях наличия или отсутствия эффектов, формализация рассуждений, формализация познавательных процедур, операции определения сходства, программная реализация приложений.

**Информационной базой** исследования являются перечисленные в списке используемой литературы научные работы по ДСМ-методу АПИ, и информация о постановке прикладных задач, полученная от экспертов предметной области.

**Основные научные результаты**, полученные в ходе исследования:

1. Разработана интеллектуальная система – компьютерный программный инструментальный интеллектуального анализа данных (ИАД), позволяющий в автоматическом и/или диалоговом режиме порождать новые знания из постоянно накапливаемых эмпирических данных (реализовать knowledge discovery).
2. При помощи развиваемого подхода и программной системы ИАД получены новые знания в клинической и молекулярно-генетической онкологии, имеющие прикладную и фундаментальную значимость, которые, согласно анализу научных источников, до настоящего времени не удавалось получить иными методами анализа данных.
3. Предложен результативный способ представления разнородных данных (генетических, количественных, клинических), используемых в онкологии, и предназначенный для ИАД, эффективность которого подтверждена экспериментами по предсказанию эффектов и получению новых знаний.
4. Разработан ДСМ-решатель задач, реализующий основные стратегии ДСМ-метода АПИ: простого сходства, сходства с запретом на контр-примеры, сходства-различия, сходства-различия с запретом на контр-примеры, а также объединение всех стратегий для повышения полноты предсказания.
5. Созданы программные инструменты, которые на языке программирования python осуществляют синтез познавательных процедур ДСМ-метода АПИ: индукции, аналогии и абдукции.
6. Экспериментально подтверждено, что созданные программные инструменты ИС результативно применимы для исследуемой предметной области (клинической и молекулярно-генетической онкологии).

7. Созданы вспомогательные программные инструменты для определения непротиворечивости массивов порожденных гипотез для всех стратегий ДСМ-рассуждений и для всех возможных последовательностей расширений.

8. Реализована функция направленной навигации в диаграмме сходств описаний прецедентов в рамках концепции приближенного ДСМ-метода, обеспечивающая в интерактивном режиме быстрый поиск интерпретируемых гипотез в исходной БФ путем направленного восстановления цепочек частичного порядка в диаграмме гипотез.

9. Разработанная интеллектуальная программная система для анализа онкологических данных обеспечена удобным графическим пользовательским интерфейсом, позволяющим не являющемуся ИТ-специалистом пользователю работать с ДСМ-решателем.

Первая ДСМ-система для анализа онкологических данных, которая, в отличие от представленной системы, не анализировала генетические данные, была создана в ФИЦ ИУ РАН О.П. Шестерниковой, средствами языка программирования С# [1]. ДСМ-решатель системы выполняется на бытовом компьютере пользователя в последовательном режиме, и поддерживает ДСМ-метод АПИ, реализующий все основные стратегии.

#### **Научная новизна** диссертационного исследования:

- реализовано автоматизированное получение новых знаний в онкологии, что имеет прикладную значимость в онкологической предметной области, как в клиническом (прогноза состояния пациента), так и в фундаментально-научном (открытие новых молекулярно-генетических механизмов возникновения и развития опухолей) направлениях;
- предложен единый способ представления генетических и прочих медицинских данных в БФ, допускающий их совместное использование в процессе ИАД. Эффективность данного способа представления данных подтверждается успешными проведенными экспериментами по ИАД для решения задач молекулярной онкологии в части прогнозирования развития



заболевания и поиска новых механизмов развития злокачественных опухолей;

- сформировано объединение всех реализованных стратегий ДСМ-метода АПИ, в результате чего достигнуто повышение полноты предсказания;
- впервые в формате ИС реализована функция направленной навигации в диаграмме сходств в рамках концепции приближенного ДСМ-метода АПИ в виде ИС, который осуществляет построение цепочек частичного порядка в диаграмме гипотез за сравнительно короткое время, что позволяет эксперту быстро проверить исходную БФ на наличие содержательных гипотез, до применения полного ДСМ-метода АПИ;
- впервые интеллектуальная система (ИС) на основе ДСМ-метода АПИ адаптирована для работы на промышленном вычислительном кластере, что дало существенное сокращение вычислительного времени, и продемонстрировало возможность запуска ДСМ-метода АПИ на высокопроизводительных суперкомпьютерах подобного класса;
- реализовано параллельное выполнение процедур ДСМ-решателя задач сразу на 3 уровнях: эффект, стратегии, перестановки расширений.

#### **Практическая значимость диссертационного исследования:**

- создана компьютерная интеллектуальная система на базе ДСМ-метода АПИ, которая позволяет пользователю (исследователю предметной области) проводить интеллектуальный анализ онкологических данных с целью прогнозирования заболевания и выявления ключевых биологических механизмов возникновения и развития злокачественной опухоли;
- система позволяет оперировать реальными исследовательскими данными, в том числе – постоянными пополнениями текущих эмпирических данных вновь получаемыми сведениями;
- реализовано прогнозирование онкологического заболевания «меланома кожи», то есть выявление пациентов, у которых ожидается наступление состояния «ремиссия» (отсутствие признаков заболевания), а также тех

пациентов, у которых ожидается наступление состояния «рецидив» (наличие признаков заболевания). Данная возможность позволяет клиническим специалистам НМИЦ онкологии им. Н.Н. Блохина выявлять пациентов с повышенным риском развития рецидива, которые классифицируются в группу усиленного мониторинга;

- выполнена классификация ранее неизвестных нарушений генов ДНК (мутаций) в соответствии с их патогенным влиянием на организм пациента: выявление среди них таких мутаций, которые влияют на характер заболевания. Данная задача важна с точки зрения определения новых прогностических факторов, необходимых для планирования стратегии лечения, а также выявления биологических механизмов онкозаболевания с целью разработки новых препаратов;
- в процессе экспериментов были выявлены нарушения генов, которые ранее не были известны в качестве вносящих весомый вклад в возникновение и развитие заболевания меланомы;
- реализована функция направленной навигации по диаграмме сходств, в рамках приближенного ДСМ-метода АПИ, позволяющая пользователю быстро находить в общем массиве гипотез такие гипотезы, которые обладают интересующими эксперта свойствами, что также позволяет проверять исходную БФ на предмет наличия интерпретируемых гипотез. Функция была протестирована экспертами НМИЦ онкологии им. Н.Н. Блохина и используется ими для решения исследовательских задач;
- в программной реализации ИС использован язык python, что в силу его широкой распространенности потенциально расширяет круг исследователей, способных проводить адаптацию ИС под собственные нужды, а также позволяет проводить несложную интеграцию с другими модулями.

Разработанная программная система может быть использована для ведения ИАД также и в ряде других значимых предметных областей при решении задач диагностического типа (в технической диагностике,

противодействии мошенничествам в банковской и финансовой сфере, обеспечении кибербезопасности и т.п.).

**Апробация результатов диссертации.** Основные положения диссертационной работы были изложены на конференции «Интеллектуальные системы в гуманитарной сфере. Опыт преподавания и развитие образовательных программ» в рамках международного научно-практического форума «Россия в XXI веке: глобальные вызовы, риски и решения» (29 мая 2019 г., РГГУ), на международной конференции «Меланома и опухоли кожи-2018» (18-19 мая 2018 г., Москва Holiday Inn Лесная), на конференции «25 лет научных исследований и образовательных программ Отделения интеллектуальных систем в гуманитарной сфере» (18 декабря 2017 г., РГГУ), на конференции «Tumor Immunology and Immunotherapy – 2020» American Association for Cancer Research (19-20 октября 2020) в докладе "Method for predicting the effectiveness of the developed immune dendritic cell vaccine in melanoma patients based on cell surface antigens and machine learning with non-classical logic", который опубликован в журнале "Cancer Immunology Research".

Созданная ИС используется в ФГБУ «НМИЦ онкологии им. Н.Н. Блохина» Минздрава РФ для решения прикладных задач при прогнозировании развития онкологических заболеваний (выявлении групп риска для применения усиленного мониторинга) а также при поиске новых биомаркеров возникновения и развития злокачественных опухолей.

**Публикации.** По теме диссертации на текущий момент опубликовано 4 статьи в сборниках Научно-техническая информация, Серия 2 (Информационные процессы и системы), рекомендованных ВАК. Все они также были переведены на английский язык и опубликованы в журнале Automatic Documentation and Mathematical Linguistic (WoS).

Отдельные результаты включались в отчеты по проектам РФФИ: № 18-29-03063 МК «Комплекс методов и средств интеллектуального анализа

слабоструктурированных данных для медицины и социологии» 2018-2021 гг., № 17-07-00539 А «Интеллектуальная система для обнаружения эмпирических закономерностей в последовательностях баз фактов» 2016-2019 гг.

**Личный вклад.** Все выносимые на защиту результаты получены автором лично при научном руководстве д.ф.-м.н. Забежайло М.И.

**Структура диссертационного исследования.** Текст диссертации состоит из введения, пяти глав, заключения, списка литературы, списка иллюстраций (рисунков и таблиц) и трех приложений, и насчитывает 121 страницу.

### **КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ**

Во **Введении** обоснована актуальность темы диссертационного исследования, определены предмет исследований, цель и задачи работы, сформулированы основные результаты, охарактеризованы научная новизна и практическая значимость представляемой работы, описана апробация результатов, дана краткая информация о публикациях и структуре текста.

В диссертации дается теоретическое и экспериментальное обоснование эффективности применения ДСМ-метода АПИ к онкологическим данным.

**Первая глава** «Обзор методов искусственного интеллекта для онкологических данных и постановка задачи» описывает актуальные задачи в сфере исследования онкологических заболеваний, используемые типы представления знаний, логико-математический и программный инструментарий, постановку задачи диссертационной работы, а также обоснование выбора метода интеллектуального анализа данных.

Современная высокотехнологичная медицина располагает огромным количеством накопленных данных, и одной из основных задач для нее на сегодня является обработка этих данных с целью получения нового знания, с учетом их адекватного формализованного представления.

Актуальными прикладными задачами являются следующие: во-первых это прогнозирование рецидива онкологического заболевания «меланома» после лечения. В этом случае у пациента возможны два исхода: ремиссия, то есть отсутствие признаков заболевания, и рецидив, то есть возвращение заболевания. Решение такой задачи очень важно в клинической практике, так как позволяет формировать группы усиленного мониторинга среди пациентов, и превентивно реагировать на проявления болезни.

Вторая задача, которая связана с первой – выявление таких патогенных мутаций, которые влияют на прогноз заболевания, то есть, иными словами, определение факторов причинности наступления одного из двух исходов (ремиссии или рецидива). Эта задача важна с точки зрения определения маркеров заболевания, и открытия новых биологических механизмов в онкологии, что может использоваться для создания новых противоопухолевых препаратов.

Еще одна задача, которая решалась в рамках представленной работы – это прогнозирование эффективности противоопухолевой вакцины, разработанной в Национальном исследовательском центре онкологии имени Блохина, а также выявление иммунных маркеров ее эффективности.

Решение перечисленных задач требует подбора средств ИАД, включая выбор подхода и его настройку, так как прямым применением традиционных подходов получить приемлемое решение названных задач пока не удается (на текущий момент нет данных о таких решениях). Необходим подход, обеспечивающий получение нового знания из уже имеющихся данных и знаний.

Знания в исследуемой предметной области плохо формализованы, так как отсутствуют универсальные эмпирические зависимости, которые бы описывали эти данные. При этом они хорошо структурированы, так как имеют описания признаков и эффектов.

Можно выделить ряд аспектов, характеризующих принципиально значимые ограничения применимости традиционных методов компьютерного анализа данных и машинного обучения в решении задач обсуждаемого типа:

- *многофакторная* причинность (взаимодействие многих причинных влияний) возникновения анализируемых эффектов и явлений;
- существенная роль *индивидуальных особенностей* анализируемых прецедентов (так в медицинской области это - индивидуальные особенности конкретных пациентов – их физического состояния, особенностей иммунной системы и др.);
- проблемы с адекватностью традиционных (в частности – для медицины) методов статистического анализа данных: часто наблюдаемы здесь эффекты больших объемов (*Big*) и открытости (*Open*) анализируемых сведений, когда данных не только много, но они постоянно пополняются новой – в том числе, принципиально новой<sup>1</sup> - информацией. Критически значимое следствие – отсутствие корректных представлений о *генеральной совокупности, выборках* из нее, *репрезентативности* таких выборок и т.п.);
- проблемы с адекватностью традиционных методов машинного обучения (*ненаследуемость* интерполяций обучающей выборки при ее *пополнении* новыми прецедентами);
- проблемы с использованием искусственных нейронных сетей: нет приемлемых *data set*'ов для их надежного обучения.

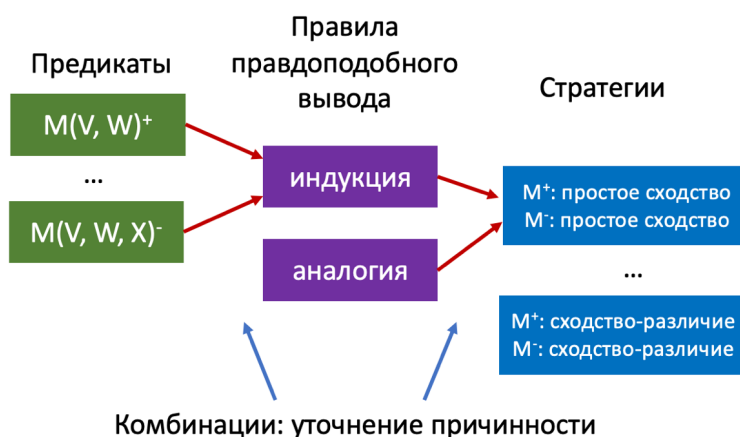
**Вторая глава** «Описание ДСМ-метода АПИ» посвящена общему описанию ДСМ-метода автоматизированной поддержки исследований, реализации ДСМ-рассуждений, и постановке проблем реализации.

Согласно процедурной конструкции ДСМ-метода, для поиска причинности используются априорно формулируемые условия, которые вначале выражаются через предикаты, то есть условия вида «при выполнении ограничений  $R$  свойство  $V$  есть причина эффекта  $W$ ». Причинность выражается с помощью сходства, уточнение представлений о котором заключается в следующем: если имеются описания двух прецедентов существования эффекта, то признаки этого эффекта, или его факторы влияния, «попадут» в локальное сходство этих

---

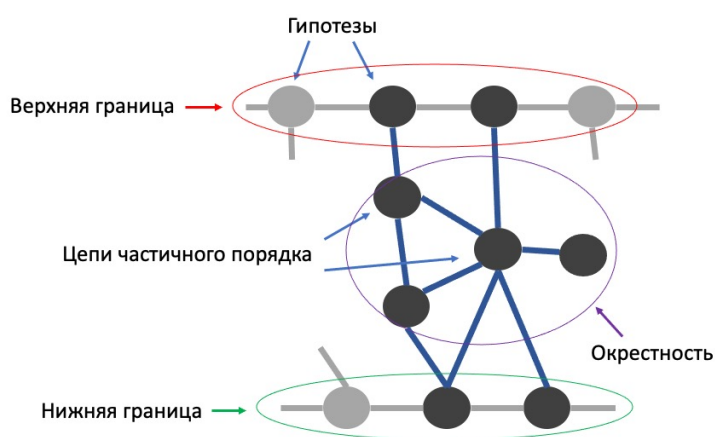
<sup>1</sup> См., в частности, проблему постоянной мутации вирусов и т.п.

двух описаний. Предикат обеспечивает определение глобального сходства примеров, то есть множества всех тех и только тех признаков, которые характеризуют эффект, у всех тех и только тех прецедентов, сходством которых является это множество признаков. Предикаты (см. Приложение на стр.27) комбинируются в правила правдоподобного вывода (ППВ), комбинации которых порождают стратегии. Все эти условия характеризуют причинность (Рис.1).



**Рисунок 1 - Отношение предикатов, ППВ и стратегий ДСМ-метода**

Порождаемые на заданной обучающей выборке сходства можно упорядочить по взаимной вложимости, определив на их множестве отношения частичного порядка, и породив диаграмму вложимости сходств (Рис.2).



**Рисунок 2 - Пример фрагмента диаграммы сходств**

Таким образом, ДСМ-метод предлагает язык для описания причинности. Эмпирические зависимости, получаемые с его помощью, есть результат причинного анализа. Ими являются объекты (комбинации признаков и комбинации примеров), которые превращают условия, представленные в предикатах,

ППВ и стратегиях, в тождества. Идентифицированные при помощи ДСМ-метода комбинации признаков являются гипотезами о причинах эффекта.

Синтез трех познавательных процедур ДСМ-метода АПИ: индукции (для анализа данных), аналогии (для предсказания изучаемых эффектов) и абдукции (для принятия порождённых гипотез с использованием объяснения) реализует ДСМ-решатель (Рис. 3.).



Рисунок 3 - Процедурная блок-схема ДСМ-метода АПИ

В зависимости от способа усиления ППВ говорят о различных стратегиях ДСМ-метода АПИ [4]. Они являются способами описания различных типов отношений причинности, и предоставляют исследователю-пользователю системы более гибкий инструментарий по сравнению с применением какой-либо одной выбранной стратегии. Стратегии являются комбинациями правил правдоподобного вывода, и за счет разных логических условий возможно строить разные правила и комбинировать их.

ДСМ-решатель реализует ДСМ-рассуждение, то есть обнаружение эмпирических закономерностей в последовательно расширяемых БФ. Последовательные расширения организуются следующим образом: исходная БФ  $A_0$ , состоящая из  $m_0$  примеров, используется для порождения гипотез посредством



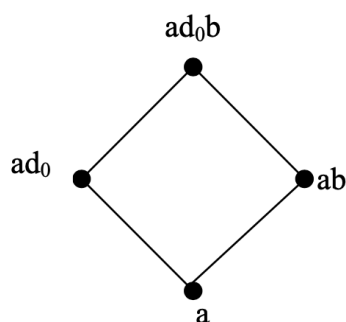
реализации ДСМ-рассуждения. Затем в  $A_0$  добавляется последующее расширение  $A_1$  из  $m_1$  элементов, такое, что  $A_0 \cap A_1 = \emptyset$ . К полученной БФ  $A_0 \cup A_1$  также применяется ДСМ-рассуждение. В процессе вычисления гипотез расширения располагаются во всех возможных последовательностях, при этом число возможных перестановок составляет  $n!$  для  $n$  расширений. Перестановки минимизируют случайность в организации БФ, и повышают тем самым надежность результата (Рис. 4).



**Рисунок 4 - Пример организации перестановок для БФ**

ДСМ-метод состоит из ДСМ-рассуждения, в результате которого порождаются гипотезы о причинах исследуемых эффектов и гипотезы о предсказаниях посредством данных причин, а также ДСМ-исследования, обеспечивающего обнаружение сохранения гипотез в последовательно расширяемых базах фактов. Результатом ДСМ-исследования является порождение знаний об эмпирических закономерностях, что квалифицируется как knowledge discovery, т.е. как цель интеллектуального анализа данных.

Стратегии помогают описывать классы эвристик, характеризующих способ рассуждения эксперта: они могут быть более простыми либо более сложными, и упорядочены в виде решетки условий, где к простому сходству добавляются дополнительные логические условия (Рис. 5).



**Рисунок 5 – Стратегии, упорядоченные в виде решетки**

В частности, такими условиями являются запрет на контр-примеры, при котором ни одна гипотеза о причинах для рассматриваемого эффекта не должна вкладываться в пример (т.е. быть подмножеством примера) противоположного эффекта, а также условие различия, когда разность между гипотезой и любым примером, участвовавшим в определении сходства, не вкладывается ни в один из примеров того же эффекта. Стратегии могут объединяться для повышения полноты предсказания.

В работе реализованы 16 стратегий, порождаемых прямым произведением четырех предикатов сходства ДСМ-метода АПИ для знаков «плюс» и «минус» ( $M^+$ -предиката и  $M^-$ -предиката): простого сходства ( $a$ ), сходства с запретом на контр-примеры ( $b$ ), упрощённый метод сходства-различия ( $d_0$ ), упрощённый метод сходства-различия с запретом на контр-примеры:

$$\text{Str}_{x,y}, x \in \{a^+, (ab^+), (ad_0^+), (ad_0b^+)\}, y \in \{a^-, (ab^-), (ad_0^-), (ad_0b^-)\}$$

Реализация ДСМ-метода АПИ связана со значительной вычислительной сложностью. С целью выбора представляющих наибольший интерес интерпретируемых гипотез и одновременного сокращения вычислительного времени был применен метод направленной навигации по множеству гипотез. Такой подход относится к концепции приближенного ДСМ-метода АПИ.

Множество гипотез образуют диаграмму сходств, в которой особый интерес представляют верхняя и нижняя границы [6]. Верхняя граница образована пересечениями минимальных подмножеств примеров, а следовательно, ее элементы имеют наибольшее количество признаков. Нижняя граница диаграммы, напротив, образована максимальными подмножествами примеров («родителей»), соответственно, число признаков у ее элементов минимально.

Направленная навигация реализуется следующим образом: эксперт-пользователь просматривает ключевые структурные элементы (нижнюю и верхнюю границы) диаграммы, после чего, при обнаружении на границах гипотез, представляющих интерес, им осуществляется движение по цепям частичного порядка в диаграмме путем последовательного перебора элементов

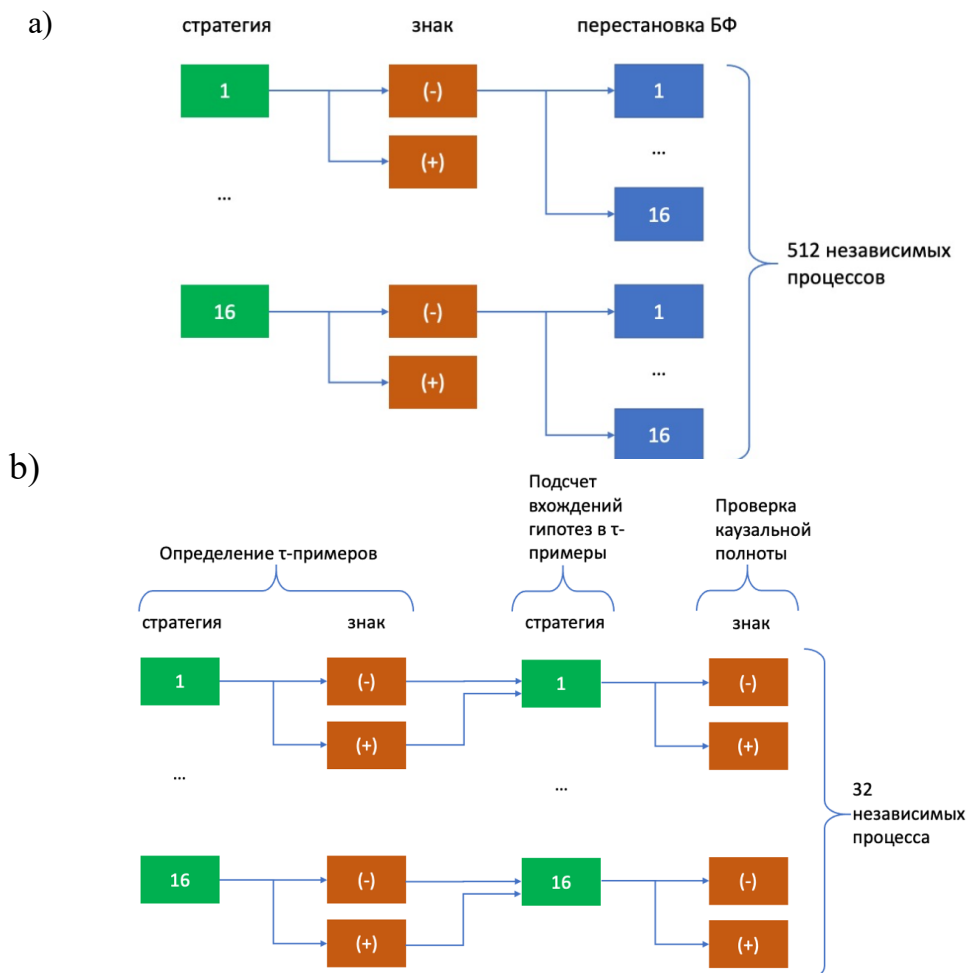
по таким цепям диаграммы от выделенных экспертом сходств верхней границы к нижней и наоборот. Каждая из таких цепей между границами состоит из вкладывающихся друг в друга гипотез, по которым можно вести движение (Рис.2). Длина каждой цепи ограничена сверху размерами множества признаков, задействованных в описаниях исходной обучающей выборки [6]. Общее время эксперимента значительно уменьшается в связи с тем, что задача по вычислению каждой отдельно взятой цепочки частичного порядка, выделяемой при движении от элемента нижней границы к элементу верхней границы или наоборот, имеет полиномиальную сложность, в отличие от доказуемо вычислительно трудной задачи формирования всей диаграммы гипотез. Размер последней в общем случае растет экспоненциально быстро с линейным ростом размеров обучающей выборки.

Для обеспечения возможности реализации ДСМ-метода АПИ за обозримое вычислительное время необходимо применять оптимизационные техники, в частности параллельное выполнение процедур, что подходит для работы с современными вычислительными машинами, обладающими многоядерной архитектурой. Вычислительные процессы в ДСМ-решателе могут выполняться независимо друг от друга на трёх различных уровнях:

- 1) обработка (+)-примеров (имеющих истинностное значение «фактически истинно») и (-)-примеров (истинностное значение «фактически ложно»);
- 2) обработка всех возможных перестановок расширений БФ: в каждом из вычислительных потоков может обрабатываться отдельная последовательность расширений БФ;
- 3) выполнение стратегий: реализация каждой из стратегий является независимым процессом, поэтому она может выполняться в отдельном потоке, соответственно для 16 стратегий возможно выделить 16 потоков, каждый из которых будет выполняться независимо от остальных.

**Третья глава** «Алгоритмическая реализация ДСМ-метода» описывает выбор способа программной реализации, параллельное выполнение процедур, оптимизацию работы с памятью, и результаты примененной оптимизации.

В качестве языка программирования был выбран *python* (версия 3.7). Выбор такого распространенного языка позволяет повысить интегрируемость системы с другими средствами подготовки и анализа данных. На Рис. 6 представлена схема организации параллельных потоков сразу на 3 уровнях.



**Рисунок 6 - Схема организации параллельных потоков на этапе порождения гипотез (а) и предсказания примеров (б)**

Процесс определения пересечений (сходств описаний) примеров является наиболее вычислительно сложной задачей в реализации ДСМ-метода АПИ. Ситуация с потреблением ресурсов памяти при реализации на языке *python* усугубляется динамической типизацией переменных. Был применен ряд приемов, связанных с особенностями анализируемых массивов данных.

В реализациях ДСМ-метода АПИ применяется ряд алгоритмов определения пересечения наборов признаков, однако одним из наиболее эффективных является алгоритм Норриса [7]. Реализован алгоритм Норриса с применением шагов по оптимизации работы с памятью (Рис. 7 и Листинги 1 и 2).

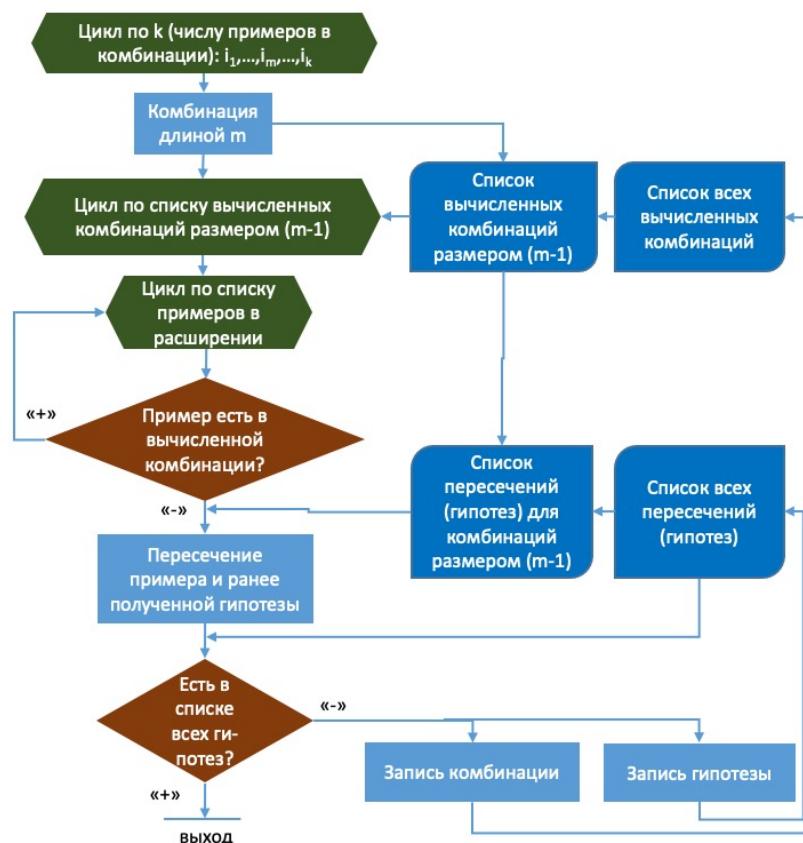


Рисунок 7 - Схема организации поиска пересечений примеров

Объявление пустого словаря (*dictionary*) для хранения гипотез и породивших их комбинаций примеров

Объявление пустого множества (*set*) для хранения всех порожденных гипотез

Цикл: Для каждого расширения в перестановке расширений БФ:

Объявление пустого множества (*set*) для хранения гипотез, порожденных в текущем расширении

Цикл: Для каждого количества пересекаемых примеров:

Если пересекаем два примера:

Цикл: Для каждого индекса примеров в текущем расширении:

Цикл: Для каждого индекса примеров во всей БФ:

Если индексы не равны и представлены в возрастающем порядке:

Нахождение пересечения

Реализация стратегии (пример – стратегия с запретом на контр-примеры):

Цикл: Для каждого примера противоположного знака:

Если порожденная гипотеза содержится в примере противоположного знака:

Выход из цикла

Иначе:

Добавить гипотезу в множество для хранения гипотез, порожденных в текущем расширении

Добавить гипотезу в множество для хранения всех порожденных гипотез

Если размер множества для хранения всех порожденных гипотез изменился (данной гипотезы в нем ранее не было):

Добавить гипотезу в словарь для хранения гипотез и породивших их комбинаций примеров, вместе с комбинацией породивших ее примеров.

Если пересекаем более чем два примера ( $m$  примеров):

Цикл: Для каждой комбинации индексов примеров, имеющих размерность  $m-1$ :

Цикл: Для каждого индекса примеров в текущем расширении:  
 Если индекс примера присутствует в ранее вычисленной комбинации:  
 Выход  
 Иначе:  
 Нахождение пересечения  
 Реализация стратегии  
 Добавить гипотезу в словарь для хранения гипотез и породивших их комбинаций примеров вместе с комбинацией породивших ее примеров

### Листинг 1 - Псевдокод модуля, осуществляющего порождение гипотез

1. Определение  $\tau$ -примеров (выполняется независимо для каждого знака):  
 Цикл: Для каждого  $\tau$ -примера в списке примеров, оставленных на предсказание:  
 Создание (очистка) рабочих файлов  
 Цикл: для каждой гипотезы о причинах:  
 Если гипотеза вкладывается в  $\tau$ -пример:  
 В файл, хранящий число вхождений гипотез в  $\tau$ -пример, в строку с номером текущего  $\tau$ -примера записывается (прибавляется) единица  
 Гипотеза записывается в файл гипотез, участвующих в предсказаниях, с ярлыком номера примера и знака
2. Подсчет вхождений гипотез в  $\tau$ -пример:  
 Определение знака примера: из файла, хранящего число вхождений гипотез в  $\tau$ -пример, считывается количество вхождений гипотез каждого знака, сравнивается между собой и делается вывод о предсказанном знаке  $\tau$ -примера.  
 Пополняется файл с перечнем (+)-примеров  
 Пополняется файл с перечнем (-)-примеров
3. Проверка каузальной полноты (выполняется независимо для каждого знака):  
 Считываются индексы примеров текущего знака  
 Из БФ считываются наборы признаков для каждого из примеров текущего знака  
 Цикл: для каждого из примеров текущего знака:  
 Цикл: для каждой из гипотез о причинах текущего знака:  
 Если гипотеза вкладывается в пример:  
 Индекс примера записывается в файл  
 Цикл по гипотезам прерывается  
 Иначе:  
 Цикл по гипотезам продолжается  
 Иначе:  
 Пример помечается как необъясненный  
 Запись в файл протокола исследования списка индексов объясненных и необъясненных примеров

### Листинг 2 - Псевдокод модуля, осуществляющего прогноз по неизвестным примерам

Представленный способ реализации ДСМ-метода АПИ был выполнен на бытовом компьютере с 6-ядерным процессором и 8 ГБ оперативной памяти, а также на вычислительном кластере Amazon AWS EC2 Instance r5.8xlarge с 32 виртуальными ядрами и 256 ГБ оперативной памяти. При запуске на бытовом компьютере параллельное выполнение заняло в 4,3 раза меньше времени, чем последовательное (9,2 часа против 39,3 соответственно).

Согласно архитектуре компьютера, максимальное число процессов, которые могли идти параллельно – шесть, что соответствует числу ядер процессора. Вычислительный кластер имел 32 ядра процессора, что подразумевает

одновременную обработку 32 процессов. При запуске на нем программы время ее выполнения при параллельном варианте оказалось меньше времени выполнения при последовательном в 21,1 раз (1,6 часа против 34,1 соотв.).

**Четвертая глава** «Описание интеллектуальной системы» описывает компоненты разработанной ИС и ее графический интерфейс.

Архитектурная схема разработанной ИС представлена на Рис.8.



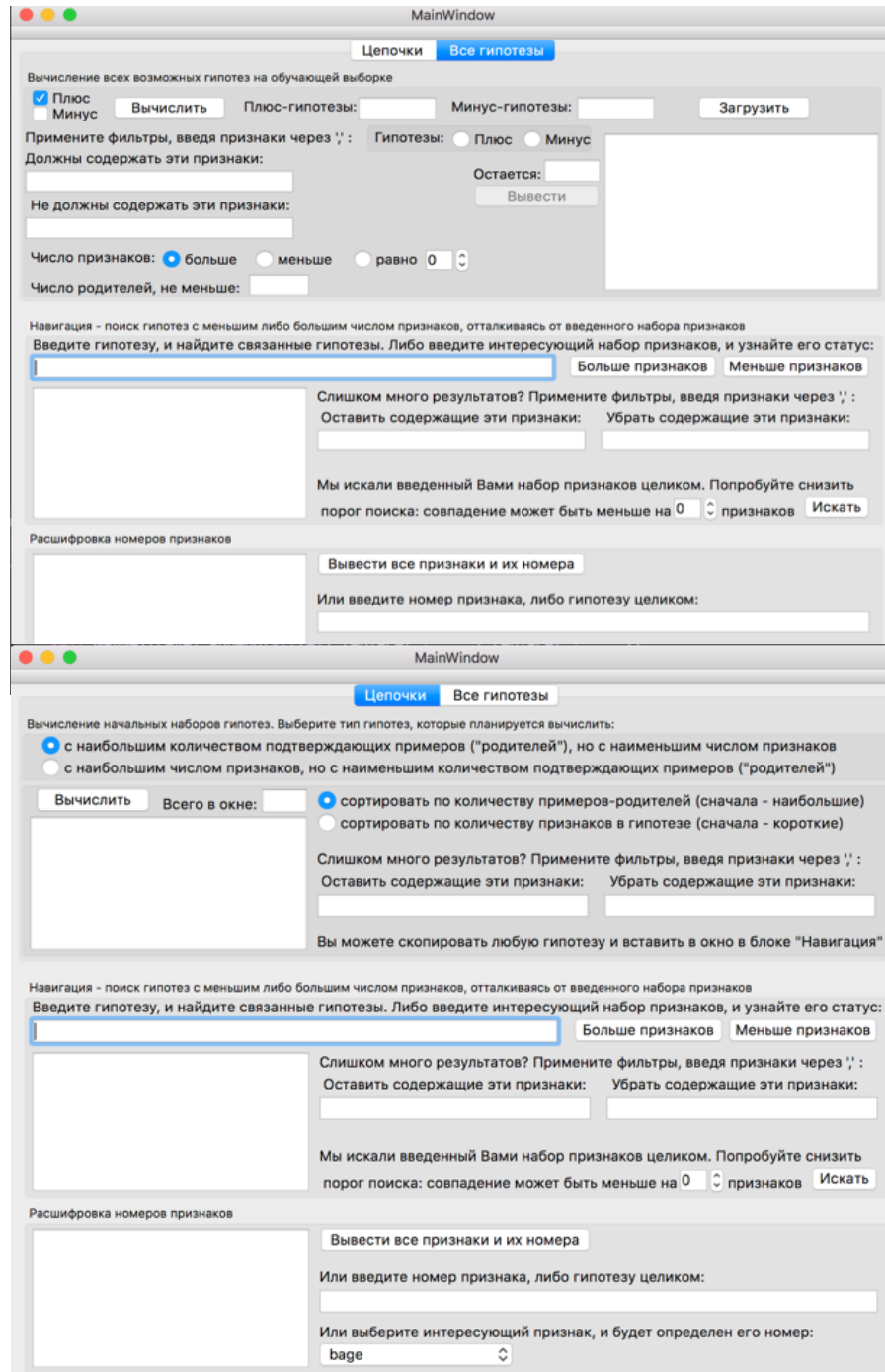
**Рисунок 8 - Архитектура интеллектуальной системы**

Пользовательский интерфейс предназначен для взаимодействия пользователя с исходной базой фактов, настройками ДСМ-решателя и результатами исследования. Пример интерфейса разработанной ИС представлен на Рис. 9.

**Пятая глава** «Эксперименты по интеллектуальному анализу онкологических данных» включает описание БФ, описывает представление клинических, генетических и иммунных знаний, постановку задачи проведения эксперимента, объединение стратегий, результаты экспериментов, а также обзор перспектив развития интеллектуальной системы, включая ее практическую значимость и универсализацию.

При решении поставленных задач эксперименты в работе проводились на данных 414-ти пациентов с генетическими и фенотипическими (клиническими и иммунными) параметрами, - в совокупности 491 признак. Из них примерно половина имела позитивный эффект, то есть состояние ремиссии не менее 2 лет, другая половина имела состояние рецидива, наступившее в течение 2 лет,

то есть негативный эффект. Для обучения было выделено 400 пациентов, остальные 14 оставлены на оценку качества обучения. Задача предсказания эффекта вакцины имела данные по иммунному статусу 39 пациентов и 21 количественный признак.



**Рисунок 9 – Пример пользовательского интерфейса ИС**

Потребовалось создание единого способа представления знаний об анализируемой предметной области. При этом необходимо обеспечить интерпре-



тируемость результатов, для чего при представлении знаний следует сохранить физический смысл исходных данных, иначе невозможно будет отсеять возможные артефакты интеллектуального анализа данных. Также имеющиеся знания необходимо оформить в виде, который позволит работать с ними при помощи предикатов, правил правдоподобного вывода и стратегий.

Исходные данные необходимо сравнивать по сходствам при помощи выделения фрагментов причины согласованным образом (однотипными средствами). Каждый из них был представлен значениями булевских переменных (к примеру, наличие/отсутствие мутации). Количественные иммунные данные также представлены бинарными значениями, обозначающими попадание в интервал. В этом случае данные представляются как булевские вектора, с которыми возможно работать однородным способом для определения сходства как алгебраической операции (Табл. 1).

	Эффект	Генотипические данные		
	Меланома: "-1", ремиссия: "1"	Признак № 1. Замена белка в гене DLL4, вызывающая замену аминокислоты: нет – "0", есть – "1"	...	Признак № 335. Замена белка в гене MOB1A, вызывающая замену аминокислоты: нет – "0", есть – "1".
Пациент 1	-1	1	...	0
Пациент 2	1	0	...	1
...	...	...	...	...
Пациент 414	1	0	...	0

	Эффект	Иммунные данные		
	Отсутствие ответа на терапию: "-1", Ответ на терапию: "1"	Признак 1. Значение CD3+ - в интервале от 31.1 до 38.4	...	Признак 126. Значение CD16_act - в интервале от 79.1 до 98.2
Пациент 1	-1	0	...	0
Пациент 2	1	0	...	1
...	...	...	...	...
Пациент 39	1	1	...	0

Таблица 1 – Пример представления исходных данных

Было проведено ДСМ-рассуждение с использованием 16 стратегий. Исследование проводилось на двух БФ: в первой использовались исключительно генотипические и фенотипические данные, во второй добавлены иммунные.

Результаты для БФ с иммунными данными сведены в Табл. 2, где *10* – правильные предсказания; *a* – критичные ошибки, при которых (+)-примеры

предсказаны как (–)-примеры и наоборот; *b* – некритичные ошибки, когда пример предсказан с эффектом «0» (истинностное значение «противоречие»); *c* – отказы от предсказания.

Наилучший результат, составляющий 10 верных предсказаний из 14, был достигнут при применении стратегии, в которой оба предиката сходства используют метод простого сходства. Добавление иммунных данных позволяет лучше работать стратегиям с методом сходства-различия: результативность таких стратегий становится наивысшей (10 верных предсказаний из 14).

Гипотезы, сохраняющиеся при всех возможных перестановках последовательных расширений, являются эмпирическими закономерностями (ЭЗК). В проведенном нами исследовании получены ЭЗК в количестве 73 для знака (–) и 139 для знака (+). При этом наиболее ценными являются такие ЭЗК, которые участвуют в правильных предсказаниях неизвестных примеров: они считаются эмпирическими законами (ЭЗ) [6]. Для БФ без иммунных данных было обнаружено 43 ЭЗ для знака (+) и 5 – для знака (–). Для БФ с иммунными данными было обнаружено 21 ЭЗ для знака (+) и 3 – для знака (–).

Стратегия	<i>l0</i>		<i>a</i>		<i>b</i>		<i>c</i>	
	+	-	+	-	+	-	+	-
Str <sub>a+, a-</sub> , Str <sub>ad0+, a-</sub> , Str <sub>a+, ad0-</sub> , Str <sub>ad0+, ad0-</sub>	6	4	1	3				
Str <sub>a+, ab-</sub> , Str <sub>ad0+, ab-</sub> , Str <sub>a+, ad0b-</sub> , Str <sub>ad0+, ad0b-</sub>	7			6				1
Str <sub>ab+, a-</sub> , Str <sub>ad0b+, a-</sub> , Str <sub>ab+, ad0-</sub> , Str <sub>ad0b+, ad0-</sub>	2	6	5	1				
Str <sub>ab+, ab-</sub> , Str <sub>ad0b+, ab-</sub>	6	2	1	3		1		1
Str <sub>ab+, ad0b-</sub> , Str <sub>ad0b+, ad0b-</sub>	6	2	1	4				1
объединение	6	5	1	2				

**Таблица 2 - Правильные предсказания и ошибки для стратегий**

В **Заключении** сформулированы основные результаты диссертационной работы, обозначены перспективы развития созданной ИС. На основании проведенной работы можно сделать следующие выводы: ДСМ-метод позволяет моделировать компьютерными средствами рассуждения эксперта, что в комплексе с представлением знаний обеспечивает реализацию получения нового знания. Именно таким путем было выполнено прогнозирование онкологического заболевания и выявление факторов причинности данного прогноза.

**Примеры решающих предикатов и правил правдоподобного вывода**

1) Решающий предикат *Простого положительного схождения*:

$$M_{a,n}^+(V,W) \Leftrightarrow \exists k \tilde{M}_{a,n}^+(V,W,k)$$

2) Структуру этого предиката  $\tilde{M}_{a,n}^+(V,W,k)$

$$\begin{aligned} \tilde{M}_{a,n}^+(V,W,k) \Leftrightarrow & \\ \Leftrightarrow \exists Z_1 \dots \exists Z_k \exists U_1 \dots \exists U_k & ((J_{(1,n)}(Z_1 \Rightarrow_1 U_1) \& \dots \& J_{(1,n)}(Z_k \Rightarrow_1 U_k) \& \\ & \& (Z_1 \cap \dots \cap Z_k = V) \& (V \neq \emptyset) \& \forall i \forall j (((i \neq j) \& (1 \leq i, j \leq k)) \rightarrow (Z_i \neq Z_j)) \& \\ & \& \forall X \forall Y ((J_{(1,n)}(X \Rightarrow_1 Y) \& (V \subset X)) \rightarrow ((W \subseteq Y) \& (W \neq \emptyset) \& (\bigvee_{i=1}^k (X = Z_i)))) \& (k \geq 2)), \end{aligned}$$

3) Решающий предикат *Простого отрицательного схождения* («симметричен» по отношению к предикату *Простого положительного схождения*  $M_{a,n}^+(V,W)$ )

$$M_{a,n}^-(V,W) \Leftrightarrow \exists k \tilde{M}_{a,n}^-(V,W,k)$$

4) Дополнительное логическое условие (пример) –

*Запрет на контр-примеры (ЗКП):*

$$M_{xb,n}^+(V,W) \Leftrightarrow M_{x,n}^+(V,W) \& (\text{ЗКП}^+), \text{ где}$$

$$(\text{ЗКП}^+) \Leftrightarrow \forall X \forall Y (((V \subset X) \& (W \subseteq Y)) \rightarrow (J_{(1,n)}(X \Rightarrow_1 Y) \vee J_{(\tau,n)}(X \Rightarrow_1 Y)))$$

5) Пример правила правдоподобного вывода:

Правило *Простого положительного схождения* (четыре подслучая)

$$(I)^+ \frac{J_{(\tau,n)}(V \Rightarrow_2 W), M_{a,n}^+(V,W) \& \neg M_{a,n}^-(V,W)}{J_{\langle 1,n+1 \rangle}(V \Rightarrow_2 W)},$$

$$(I)^- \frac{J_{(\tau,n)}(V \Rightarrow_2 W), \neg M_{a,n}^+(V,W) \& M_{a,n}^-(V,W)}{J_{\langle -1,n+1 \rangle}(V \Rightarrow_2 W)},$$

$$(I)^0 \frac{J_{(\tau,n)}(V \Rightarrow_2 W), M_{a,n}^+(V,W) \& M_{a,n}^-(V,W)}{J_{\langle 0,n+1 \rangle}(V \Rightarrow_2 W)},$$

$$(I)^\tau \frac{J_{(\tau,n)}(V \Rightarrow_2 W), \neg M_{a,n}^+(V,W) \& \neg M_{a,n}^-(V,W)}{J_{(\tau,n+1)}(V \Rightarrow_2 W)}.$$

## Основные результаты диссертации, выносимые на защиту:

1. Разработана интеллектуальная система – компьютерный программный инструментальный интеллектуального анализа данных (ИАД), позволяющий в автоматическом и/или диалоговом режиме порождать новые знания из постоянно накапливаемых эмпирических данных (реализовать *knowledge discovery*).
2. При помощи развиваемого подхода и программной системы ИАД получены новые знания в клинической и молекулярно-генетической онкологии, имеющие прикладную и фундаментальную значимость, которые, согласно анализу научных источников, до настоящего времени не удавалось получить иными методами анализа данных.
3. Предложен результативный способ представления разнородных данных (генетических, количественных, клинических и прочих медицинских), используемых в онкологии, и предназначенный для ИАД, эффективность которого подтверждена экспериментами по предсказанию эффектов и получению новых знаний.
4. Разработан ДСМ-решатель задач, реализующий основные стратегии ДСМ-метода АПИ: методы простого сходства, сходства с запретом на контр-примеры, сходства-различия, сходства-различия с запретом на контр-примеры, а также их объединение для повышения полноты предсказания.
5. Созданы программные инструменты на языке программирования *python*, которые осуществляют синтез познавательных процедур ДСМ-метода АПИ: индукции, аналогии и абдукции.
6. Экспериментально подтверждено, что созданные программные инструменты ИС результативно применимы для исследуемой предметной области (клинической и молекулярно-генетической онкологии).
7. Созданы вспомогательные программные инструменты для определения непротиворечивости множеств порожденных гипотез для всех стратегий ДСМ-рассуждений и для всех возможных последовательностей расширений анализируемых баз фактов.
8. Реализована функция направленной навигации в диаграмме сходств описаний прецедентов в рамках концепции приближенного ДСМ-метода, обеспечивающая в интерактивном режиме быстрый поиск интерпретируемых гипотез в исходной БФ путем направленного восстановления цепочек частичного порядка в диаграмме гипотез.
9. Разработанная ИС для анализа онкологических данных обеспечена удобным графическим пользовательским интерфейсом, позволяющим не являющемуся ИТ-специалистом пользователю работать с ДСМ-решателем.

## Публикации автора по теме диссертации:

- 1) *Чебанов Д.К., Михайлова И.Н.* Интеллектуальный анализ данных пациентов с меланомой для поиска маркеров заболевания и значимых генов // Научно-техническая информация. Сер. 2. – 2019. – № 10 – С. 35-40;

*D. K. Chebanov and I. N. Mikhaylova.* Intellectual Mining of Patient Data with Melanoma for Identification of Disease Markers and Critical Genes // *Automatic Documentation and Mathematical Linguistics*, 2019, Vol. 53, No. 5, pp. 283–287.

- 2) *Чебанов Д.К., Михайлова И.Н.* Интеллектуальная система для анализа онкологических данных, реализующая ДСМ-метод автоматизированной поддержки исследований // Научно-техническая информация. Сер. 2. – 2020. – № 5 – С. 19-29;

*D. K. Chebanov and I. N. Mikhaylova.* An Intelligent System for Analyzing Oncological Data that Implements the JSM Method for Automated Research Support // *Autom. Docum. and Math. Linguistics*, 2020, Vol. 54, No. 3, pp. 130–139.

- 3) *Чебанов Д.К.* Об особенностях реализации ДСМ-метода для интеллектуального анализа данных // Научно-техническая информация. Сер. 2. – 2020. – № 7 – С. 21-26;

*D. K. Chebanov.* On the Features of Implementation of the Solver of the JSM Method for Intellectual Data Analysis // *Autom. Docum. and Math. Linguistics*, 2020, Vol. 54, No. 4, pp. 196–201.

- 4) *Чебанов Д.К., Михайлова И.Н.* О методах искусственного интеллекта для анализа онкологических данных // Научно-техническая информация. Сер. 2. – 2020. – № 9 – С. 21-26;

*D. K. Chebanov and I. N. Mikhaylova.* On the Methods of Artificial Intelligence for Analysis of Oncological Data // *Autom. Docum. and Math. Linguistics*, 2020, Vol. 54, No. 5, pp. 255–259.

- 5) *Chebanov DK, Mikhailova IN, Tatevosova NS.* Method for predicting the effectiveness of the developed immune dendritic cell vaccine in melanoma patients based on cell surface antigens and machine learning with non-classical logic [abstract]. In: Abstracts: AACR Virtual Special Conference: Tumor Immunology and Immunotherapy; 2020 Oct 19-20. Philadelphia (PA): AACR; *Cancer Immunol Res* 2021; 9 (2 Suppl): Abstract nr PO086.

## Список использованной литературы:

1. Шестерникова О.П., Финн В.К., Винокурова Л.В., Лесько К.А., Варварина Г.Г., Тюляева Е.Ю. Интеллектуальная система для диагностики заболеваний поджелудочной железы // Научно-техническая информация. Сер. 2. – 2019. – № 10. – С. 41-48.
2. Михайлова И. Н., Панкратова Е.С., Добрынин Д.А., Самойленко И.В., Решетникова В.В., Шелепова В.М., Демидов Л.В., Барышников А.Ю., Финн В.К. О применении интеллектуальной компьютерной системы для анализа клинических данных больных меланомой. // Российский Биотерапевтический Журнал. – 2010. – № 2. – С.54
3. Финн В.К. Дистрибутивные решетки индуктивных ДСМ-процедур // Научно-техническая информация. Сер. 2. – 2014. – № 11. – С. 1-36.
4. ДСМ-метод автоматического порождения гипотез: Логические и эпистемологические основания / сост. О.М. Аншаков, Е.Ф. Фабрикантова; под общ. ред. О.М. Аншакова. – М.: ЛИБРОКОМ, 2009. – 433 с.
5. Финн В.К. Об определении эмпирических закономерностей посредством ДСМ - метода автоматического порождения гипотез // Искусственный интеллект и принятие решений. – 2010. – № 4. – С. 41-48.
6. Забежайло М.И. Приближенный ДСМ-метод на примерах // Научно-техническая информация. Сер.2. – 2014. – №10. – С.1-12.
7. Финн В.К., Шестерникова О.П. О новом варианте обобщенного ДСМ-метода автоматизированной поддержки научных исследований // Искусственный интеллект и принятие решений. – 2016. – № 1. – С. 57-64