
ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ МОДЕЛИРОВАНИЯ И УПРАВЛЕНИЯ

НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

Основан в 1994 г.

**2026
№ 3(145)**

Воронеж

Издательство «Научная книга»



2026

Издательство "Научная книга"
Липецкий государственный технический университет
Бакинский государственный университет

ISSN 1813-9744

Журнал зарегистрирован в Центрально-Черноземном управлении Федеральной службы по надзору за соблюдением законодательства в сфере массовых коммуникаций и охране культурного наследия

П И Н Ф С 6-0238 от 19 сентября 2005 г.

Журнал выходит один раз в квартал

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ МОДЕЛИРОВАНИЯ И УПРАВЛЕНИЯ

РЕДАКЦИОННАЯ КОЛЛЕГИЯ:

Главный редактор О.Я.Кравец, д-р техн. наук, профессор

ЧЛЕНЫ РЕДАКЦИОННОЙ КОЛЛЕГИИ:

А.А.Алиев, д-р техн. наук, профессор (БГУ, Азербайджан, г. Баку)

С.Л.Блюмин, д-р физ.-мат. наук, профессор (ЛГТУ, Россия, г. Липецк)

С.Л.Подвальный, д-р техн. наук, профессор (ВГТУ, Россия, г. Воронеж)

Статьи, поступающие в редакцию, рецензируются. За достоверность сведений, изложенных в статьях, ответственность несут авторы публикаций. Мнение редакции может не совпадать с мнением авторов материалов. При перепечатке ссылка на журнал обязательна.

Правила для авторов доступны на сайте журнала <http://www.sbook.ru/itmu>

Материалы публикуются в авторской редакции.

Дизайн обложки – С.А.Кравец

Адрес редакции и издателя:
394077 Воронеж, ул. 60-й Армии, дом 25,
комн. 120

Телефон: (473)2667653
E-mail: itmu@yandex.ru
<http://www.sbook.ru/itmu>



Подписной индекс в объединенном каталоге «Пресса России» - 41932

Учредитель и издатель: ООО Издательство "Научная книга", <http://www.sbook.ru>

Отпечатано с готового оригинал-макета в ООО "Цифровая полиграфия"

Адрес типографии: 394036, г.Воронеж, ул. Куколкина, 6, тел.: (473) 261-03-61

Подписано в печать 01.09.2026.

Свободная цена

Заказ 0000. Тираж 1000. Усл. печ. л. 5,3. Дата выхода в свет 09.10.2026.

Содержание

1. Информационные технологии в приложениях

Кравец О.Я. Экономика, управление и здоровье: обзор актуальных публикаций	168
Немцов Л.Б., Немцов М.Л. Сквозная наблюдаемость ERP-систем	182

2. Моделирование и анализ сложных систем

Мадетбаева Н.Б. Гибридная модель Байеса и градиентного бустинга на основе онтологии узбекских фразеологизмов	192
Назирова Э., Бахромова Л. Экономико-математическая модель определения численности официального и неформального персонала	203

3. Программные и телекоммуникационные системы

Беспалов П.В. Управление гетерогенными данными в больших системах ..	215
Кравец О.Я. Проблема консенсуса, инициируемого событиями, для FOMAS в нечеткой топологии Такаги – Сугено	232

1. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ПРИЛОЖЕНИЯХ

Кравец О.Я.

ЭКОНОМИКА, УПРАВЛЕНИЕ И ЗДОРОВЬЕ: ОБЗОР АКТУАЛЬНЫХ ПУБЛИКАЦИЙ

Воронежский государственный технический университет

Настоящее исследование проведено на базе журналов «Экономика и менеджмент систем управления (2023)» и «Журнал экспериментальной, клинической и профилактической медицины (2022, 2023)».

В рамках современных акушерско-гинекологических исследований особое внимание уделяется изучению беременности как триггерного фактора, способного индуцировать реактивацию латентных форм герпесвирусной инфекции. В работе [1] представлены результаты комплексного анализа, посвящённого оценке прогностических возможностей в отношении активации латентной герпесвирусной инфекции в период гестации. Авторы рассматривают иммунологические и вирусологические механизмы, лежащие в основе реактивации, а также предлагают методологические подходы к построению прогностических моделей, что имеет существенное значение для стратификации рисков и своевременной профилактики перинатальных осложнений.

Эпидемиологические аспекты онкологической патологии нашли отражение в исследовании [2], где представлены данные о заболеваемости и смертности от трижды негативной (triple-negative) формы рака молочной железы на территории Апшеронского полуострова Азербайджанской Республики. Указанная форма характеризуется отсутствием экспрессии рецепторов эстрогена, прогестерона и HER2/неу, что определяет её агрессивный клинический фенотип и ограниченные терапевтические опции. Проведённый анализ позволяет выявить региональные особенности распространённости данной патологии и обосновать необходимость адаптации скрининговых программ с учётом территориальной специфики.

В контексте репродуктивной медицины значительный интерес представляет работа [3], в которой детально описаны принципы прегравидарной подготовки пациенток, страдающих привычным невынашиванием беременности на фоне хронического эндометрита. Данный подход базируется на комплексной оценке состояния эндометрия, включающей морфологические, иммуногистохимические и микробиологические методы исследования, и предполагает персонализированную коррекцию выявленных нарушений на этапе, предшествующем зачатию. Логическим следствием такого подхода является снижение частоты репродуктивных потерь и улучшение перинатальных исходов.

Теоретические основы управления стохастическими системами рассматриваются в исследовании [4], посвящённом оперативному управлению вероятностными процессами в многостадийной специализированной системе. Методологическая база работы опирается на аппарат теории массового об-

служивания и марковских процессов, что позволяет формализовать процедуры принятия решений в условиях неопределённости. Полученные результаты имеют прикладное значение для оптимизации функционирования сложных технических и организационных систем.

Продолжением данной проблематики является работа [5], в которой представлены аналитические основы управления потоками системы оперативного оповещения, обладающей нестационарной структурой. Нестационарность рассматривается как фундаментальное свойство системы, требующее применения адаптивных алгоритмов управления. Авторы предлагают формализованные критерии эффективности и демонстрируют возможности аналитического моделирования потоков в условиях изменяющейся топологии системы.

В области кардиологии заслуживает внимания исследование [6], посвящённое интервенционному лечению пациентов с фибрилляцией предсердий. Особый акцент сделан на оценке влияния восстановления и удержания синусового ритма на диастолическую функцию сердца. Диастолическая дисфункция рассматривается как самостоятельный предиктор неблагоприятных исходов, что обосновывает необходимость её учёта при выборе стратегии интервенционного вмешательства. Полученные данные расширяют представления о патофизиологических механизмах ремоделирования миокарда при фибрилляции предсердий.

Существенный вклад в решение проблемы хронического эндометрита вносит работа [7], в которой доказано улучшение результатов лечения пациенток репродуктивного возраста на основе разработанного и патогенетически обоснованного комплексного подхода. Данный подход предполагает внутриматочное применение углеродного формованного сорбента, модифицированного поливинилпирролидоном. Модификация сорбента обеспечивает улучшение его биосовместимости и сорбционной ёмкости, что подтверждается результатами клинических и лабораторных исследований. Логическим следствием внедрения данной методики является повышение эффективности восстановления рецептивности эндометрия.

Молекулярно-генетические аспекты гинекологической патологии освещены в работе [8], где представлены результаты изучения экспрессии генов рецепторов стероидных гормонов в мононуклеарных клетках периферической крови пациенток с миомой матки. Выбор мононуклеарных клеток периферической крови в качестве объекта исследования обусловлен их доступностью и возможностью оценки системных гормональных эффектов. Полученные данные позволяют расширить представления о патогенезе миомы матки и открывают перспективы для разработки малоинвазивных диагностических подходов.

В инфекционной патологии значимым является исследование [9], посвящённое прогностическим критериям развития различных вариантов течения ВЭБ-инфекционного мононуклеоза, а также оценке эффективности иммунокорректирующей терапии. Вирус Эпштейна – Барр рассматривается как

один из наиболее распространённых герпесвирусов, способных вызывать широкий спектр клинических проявлений. Авторы обосновывают дифференцированный подход к терапии в зависимости от прогнозируемого варианта течения заболевания, что соответствует принципам персонализированной медицины.

Коморбидная патология внутренних органов рассматривается в работе [10], где изучены особенности артериальной ригидности у пациентов с хронической обструктивной болезнью лёгких и хронической сердечной недостаточностью. Артериальная ригидность оценивается как интегральный показатель сосудистого ремоделирования, отражающий совокупное влияние системного воспаления и гемодинамических нарушений. Полученные результаты имеют значение для стратификации кардиоваскулярного риска у данной категории пациентов.

В сфере экономики и управления исследование [11] посвящено особенностям экономических рисков, возникающих при серийных закупках техники радиоэлектронной борьбы. Специфика данной сферы обусловлена высокой стоимостью продукции, длительными сроками производства и ограниченным числом поставщиков. Авторы предлагают классификацию рисков и методы их количественной оценки, что позволяет повысить обоснованность управленческих решений в области государственных закупок.

Клинико-морфологические аспекты хронического панкреатита рассматриваются в работе [12], где описаны изменения толстой кишки и их клинко-прогностическое значение. Взаимосвязь между патологией поджелудочной железы и состоянием толстой кишки объясняется общностью эмбрионального происхождения, особенностями кровоснабжения и лимфооттока, а также системными метаболическими нарушениями. Выявленные закономерности позволяют расширить диагностический алгоритм и улучшить прогнозирование течения заболевания.

Неонатальная офтальмология представлена в исследовании [13], посвящённом этиологическим особенностям дакриоцистита новорождённых. Данное заболевание характеризуется нарушением проходимости носослёзного протока и развитием воспалительного процесса в слёзном мешке. Авторы систематизируют этиологические факторы и предлагают дифференцированные подходы к терапии с учётом выявленных особенностей.

Реабилитационные аспекты хирургического лечения рассматриваются в работе [14], где доказана роль общей магнитотерапии в реабилитации пациентов после оперативного лечения грыж межпозвонковых дисков пояснично-крестцового отдела позвоночника. Магнитотерапия рассматривается как физиотерапевтический метод, обладающий противовоспалительным, анальгетическим и регенеративным действием. Включение данного метода в комплекс реабилитационных мероприятий позволяет ускорить восстановление и снизить риск послеоперационных осложнений.

В области гастроэнтерологии заслуживает внимания работа [15], в которой представлены первые результаты радиочастотной абляции при лечении

пищевода Барретта в краевой клинической больнице №2. Пищевод Барретта рассматривается как предраковое состояние, характеризующееся метаплазией эпителия. Радиочастотная абляция позиционируется как малоинвазивный метод, позволяющий достичь эрадикации метапластического эпителия с минимальным риском осложнений.

Хирургическая колопроктология представлена в исследовании [16], посвящённом вакуум-терапии при хирургическом лечении эпителиального копчикового хода. Вакуум-терапия рассматривается как метод, обеспечивающий активное дренирование раны и стимуляцию репаративных процессов. Применение данного метода позволяет сократить сроки заживления и снизить частоту рецидивов.

Эпидемиология онкологических заболеваний отражена в работе [17], где изучены заболеваемость и смертность от рака лёгкого в Азербайджанской Республике. Рак лёгкого остаётся одной из ведущих причин онкологической смертности в мире, что обуславливает актуальность региональных эпидемиологических исследований. Полученные данные позволяют оценить эффективность действующих профилактических программ и обосновать направления их совершенствования.

Экономическая теория представлена в исследовании [18], посвящённом анализу влияния прогрессивности подоходного налогообложения на динамику экономических процессов. Прогрессивное налогообложение рассматривается как инструмент перераспределения доходов и стимулирования экономической активности. Авторы применяют эконометрические методы для оценки количественных эффектов налоговой политики.

Гигиенические аспекты профилактики рассматриваются в работе [19], где представлено гигиеническое обоснование первичной профилактики анемий, связанных с питанием. Алиментарные анемии рассматриваются как следствие дефицита железа, фолатов и витамина В12. Авторы обосновывают необходимость системных мероприятий на популяционном уровне, включающих обогащение пищевых продуктов и образовательные программы.

Педиатрическая инфектология представлена в исследовании [20], посвящённом течению, исходам и особенностям ведения ротавирусной инфекции у детей с атопическим дерматитом в острый и восстановительный периоды. Атопический дерматит рассматривается как фактор, модифицирующий течение инфекционного процесса. Полученные данные позволяют оптимизировать терапевтические подходы с учётом коморбидного фона.

В сфере сервисной экономики работа [21] посвящена автоматизированной системе рейтинговой оценки сервисных услуг фирмы. Система базируется на многокритериальной оценке и использовании методов квалиметрии. Автоматизация процесса позволяет повысить объективность и оперативность оценки качества услуг.

Ортопедическая патология детского возраста рассматривается в работе [22], где проведён анализ отдалённых результатов коррекции врождённой косялости после использования подходов Виленского-Зацепина и Понсети у

детей раннего возраста. Сравнительный анализ демонстрирует преимущества метода Понсети, заключающиеся в меньшей инвазивности и лучших функциональных исходах. Полученные результаты имеют значение для выбора тактики лечения.

Хирургическая панкреатология представлена в исследовании [23], где доказано улучшение диагностики, непосредственных и отдалённых результатов хирургического лечения больных билиарным панкреатитом. Билиарный панкреатит рассматривается как форма острого панкреатита, обусловленная патологией желчевыводящих путей. Предложенные диагностические и лечебные алгоритмы позволяют улучшить прогноз заболевания.

Челюстно-лицевая хирургия представлена в работе [24], посвящённой зависимости тактики лечения травматических повреждений челюстно-лицевой области от характера травмы. Характер травмы рассматривается как системообразующий фактор, определяющий выбор хирургической тактики. Авторы предлагают классификацию повреждений и соответствующие алгоритмы лечения.

Региональная экономика представлена в исследовании [25], где проведены оценки показателей социально-экономического развития регионов на основе теории латентных переменных. Теория латентных переменных позволяет формализовать интегральные показатели, недоступные прямому измерению. Применение данного подхода повышает обоснованность межрегиональных сравнений.

Онкоэпидемиология представлена в работе [26], где проведён ситуационный анализ заболеваемости раком яичников в Азербайджанской Республике. Рак яичников характеризуется поздней диагностикой и высокой летальностью, что обуславливает актуальность эпидемиологического мониторинга. Полученные данные позволяют оценить динамику заболеваемости и обосновать профилактические меры.

Лучевая диагностика представлена в исследовании [27], где дана оценка распространённости предикторов субхондрального стрессового перелома от недостаточности мышечков коленного сустава по данным магнитно-резонансной томографии. Магнитно-резонансная томография рассматривается как метод выбора для выявления ранних признаков стрессовых повреждений. Выявленные предикторы позволяют своевременно корректировать нагрузочный режим.

Политология и международные отношения представлены в работе [28], где описано современное состояние глобальных конфликтов в координатах «Восток – Запад» и «Север – Юг». Данные координаты рассматриваются как аналитические оси, позволяющие структурировать геополитические противоречия. Авторы анализируют динамику конфликтов и факторы, определяющие их эскалацию.

Продолжением данной проблематики является исследование [29], посвящённое экономическим и политическим аспектам незападных центров силы в управлении современной постбиполярной системой международных

отношений. Незападные центры силы рассматриваются как новые субъекты глобального управления, изменяющие конфигурацию международных отношений. Авторы анализируют их роль и потенциал влияния на формирование нового мирового порядка.

Финансовая экономика представлена в работе [30], где представлены международные тренды процесса цифровизации финансового рынка России. Цифровизация рассматривается как трансформационный процесс, затрагивающий все сегменты финансового рынка. Авторы систематизируют международный опыт и оценивают возможности его адаптации к российским условиям.

Кардиология представлена в исследовании [31], где изучены особенности ишемии миокарда у пациентов с гипертрофической кардиомиопатией. Ишемия миокарда при данной патологии может развиваться даже при отсутствии атеросклеротического поражения коронарных артерий, что обусловлено несоответствием между потребностью миокарда в кислороде и возможностями коронарного кровотока. Полученные данные имеют значение для диагностики и лечения.

Вирусология представлена в работе [32], где представлены клинико-лабораторные критерии острой цитомегаловирусной инфекции и результативность терапии у взрослых неиммунокомпрометированных больных. Цитомегаловирусная инфекция у иммунокомпетентных лиц рассматривается как самостоятельная клиническая проблема, требующая дифференцированного подхода к диагностике и лечению. Авторы предлагают критерии оценки тяжести и эффективности терапии.

Транспортная экономика представлена в исследовании [33], где показано регулирование тарифов пассажирских перевозок в условиях современных вызовов. Тарифное регулирование рассматривается как инструмент балансирования интересов перевозчиков и пассажиров. Авторы анализируют современные вызовы и предлагают адаптивные модели тарифообразования.

Перинатальная психология представлена в работе [34], где описан патогенез эмоциональных нарушений при беременности с гипертензивными расстройствами. Эмоциональные нарушения рассматриваются как следствие комплексного взаимодействия гормональных, гемодинамических и психосоциальных факторов. Полученные данные обосновывают необходимость психопрофилактического сопровождения беременности.

Хирургическая гастроэнтерология представлена в исследовании [35], где представлено лечение деструктивного панкреатита с учётом энтерального зондового питания и селективной деконтаминации кишечника. Энтеральное зондовое питание рассматривается как метод нутритивной поддержки, способствующий сохранению барьерной функции кишечника. Селективная деконтаминация направлена на снижение транслокации бактерий и профилактики инфекционных осложнений.

Информационные технологии представлены в работе [36], где исследованы особенности работы с упреждающими данными. Упреждающие данные

рассматриваются как информация, получаемая до наступления события и позволяющая осуществлять проактивное управление. Авторы предлагают методы обработки и интерпретации таких данных.

Нефрология представлена в исследовании [37], где изучены информативные лабораторные показатели прогнозирования трансформации острого гломерулонефрита в хронический на примере постинфекционного гломерулонефрита. Трансформация острого процесса в хронический рассматривается как критический этап, определяющий долгосрочный прогноз. Выявленные лабораторные показатели позволяют своевременно идентифицировать пациентов группы риска.

Кардиология представлена в работе [38], где исследована роль анизоцитоза в сравнении с классическими факторами риска и маркерами воспаления в оценке прогноза заболевания при декомпенсации хронической сердечной недостаточности. Анизоцитоз эритроцитов рассматривается как интегральный показатель, отражающий гетерогенность популяции эритроцитов и косвенно характеризующий системные нарушения. Полученные данные расширяют прогностические возможности.

Акушерство представлено в исследовании [39], где проведено клинкоморфологическое обоснование применения антикоагулянтной терапии в профилактике плацентарной недостаточности. Плацентарная недостаточность рассматривается как следствие нарушений микроциркуляции и тромбофилических состояний. Антикоагулянтная терапия позиционируется как патогенетически обоснованный метод профилактики.

Экономика труда представлена в работе [40], где представлена математическая модель влияния инвестиций в когнитивные технологии на рост производительности труда. Когнитивные технологии рассматриваются как фактор, повышающий эффективность интеллектуального труда. Математическая модель позволяет количественно оценить вклад инвестиций в рост производительности.

Эпидемиология инфекционных заболеваний представлена в исследовании [41], где показаны особенности распространения и течения инфекций, вызванных вирусами гепатитов В и С, среди онкологических больных, проживающих в Азербайджане. Онкологические больные рассматриваются как группа повышенного риска инфицирования и тяжёлого течения вирусных гепатитов. Полученные данные обосновывают необходимость скрининга и профилактических мероприятий.

Маркетинг представлен в работе [42], где обоснован момент возникновения потребности в товаре как момент возникновения вероятности боя. Данный подход позволяет формализовать процесс управления товарными запасами с учётом вероятностных характеристик. Полученные результаты имеют значение для оптимизации логистических процессов.

Туризм представлен в исследовании [43], где описан экологический туризм как перспективное направление развития туристской отрасли Алтайского края. Экологический туризм рассматривается как форма устойчивого ту-

ризма, ориентированная на сохранение природных ресурсов. Авторы анализируют потенциал региона и предлагают направления развития.

Экономическая безопасность представлена в работе [44], где представлено обеспечение экономической безопасности предприятий в условиях ограничений в сфере внешнеэкономической деятельности. Ограничения рассматриваются как фактор, требующий адаптации стратегий управления. Авторы предлагают комплекс мер по обеспечению экономической безопасности.

Офтальмология представлена в исследовании [45], где исследован эндонозальный электрофорез дексаметазона в комплексном лечении увеитов при болезни Бехтерева. Эндонозальный электрофорез рассматривается как метод, обеспечивающий доставку лекарственного средства непосредственно к очагу воспаления. Полученные данные обосновывают включение данного метода в комплексную терапию.

Хирургия представлена в работе [46], где доказано улучшение результатов лечения больных с острым аппендицитом путём оптимизации техники лапароскопической аппендэктомии. Лапароскопическая аппендэктомия рассматривается как метод выбора при остром аппендиците. Оптимизация техники позволяет снизить частоту осложнений и сократить сроки госпитализации.

Неврология представлена в исследовании [47], где проведено исследование влияния наследственности на характер эпилепсии и эпилептоформных аномалий у детей. Наследственность рассматривается как один из ключевых факторов, определяющих клиническую гетерогенность эпилепсии. Полученные данные имеют значение для генетического консультирования.

Региональное управление представлено в работе [48], где описано управление региональным развитием с учётом межрегиональных диспропорций. Межрегиональные диспропорции рассматриваются как фактор, требующий дифференцированного подхода к управлению. Авторы предлагают методы оценки и инструменты выравнивания.

Неврология представлена в исследовании [49], где проведена оценка эффективности применения ноотропов при лечении в остром периоде сотрясения головного мозга. Ноотропы рассматриваются как средства, улучшающие когнитивные функции и ускоряющие восстановление. Полученные данные обосновывают целесообразность их применения.

Офтальмология представлена в работе [50], где описаны ранние стадии диабетической оптической нейропатии с акцентом на патогенетические и клиничко-диагностические аспекты. Диабетическая оптическая нейропатия рассматривается как осложнение сахарного диабета, приводящее к снижению зрения. Ранняя диагностика имеет ключевое значение для предотвращения необратимых изменений.

Экономика представлена в исследовании [51], где проведена оценка потенциала механизмов государственно-частного партнёрства. Государственно-частное партнёрство рассматривается как институциональный механизм, по-

звolyающий привлекать частные ресурсы для решения общественно значимых задач. Авторы анализируют потенциал и ограничения данного механизма.

Онкохирургия представлена в работе [52], где представлены современные подходы к выполнению спленэктомии на этапе хирургического лечения рака желудка. Спленэктомия рассматривается как компонент расширенного хирургического вмешательства. Авторы анализируют показания и технические аспекты выполнения данной операции.

Продолжением данной проблематики является исследование [53], где проведён сравнительный анализ и определены пути улучшения результатов гастрэктомии со спленэктомией и без спленэктомии у больных раком желудка. Сравнительный анализ позволяет оценить влияние спленэктомии на непосредственные и отдалённые результаты лечения. Полученные данные имеют значение для выбора хирургической тактики.

Эпидемиология представлена в работе [54], где дан анализ заболеваемости лихорадкой Западного Нила в Российской Федерации и мире. Лихорадка Западного Нила рассматривается как трансмиссивная инфекция, распространение которой связано с климатическими изменениями. Полученные данные позволяют оценить динамику и обосновать профилактические меры.

Дерматология представлена в исследовании [55], где представлена оценка морфологии кожи у больных с хроническими дерматозами на основе ультразвукового сканирования. Ультразвуковое сканирование рассматривается как неинвазивный метод, позволяющий оценить структурные изменения кожи. Полученные данные расширяют диагностические возможности.

Транспортная логистика представлена в работе [56], где описаны методические аспекты организации деятельности в сфере оказания транспортных услуг. Организация деятельности рассматривается как комплекс управленческих решений, направленных на повышение эффективности. Авторы предлагают методические подходы к организации.

Педиатрия представлена в исследовании [57], где представлена проблема идентификации цитомегаловирусной инфекции и её влияния на формирование соматической патологии у детей. Цитомегаловирусная инфекция рассматривается как фактор, способствующий формированию хронических заболеваний. Полученные данные обосновывают необходимость ранней диагностики.

Косметология представлена в работе [58], где проведено обоснование применения и подходов к терапии локальных жировых отложений с помощью аппаратных методов – криолиполиза и радиоволнового воздействия. Криолиполиз и радиоволновое воздействие рассматриваются как неинвазивные методы, обеспечивающие селективное разрушение адипоцитов. Авторы обосновывают показания и методику применения.

Продолжением данной проблематики является исследование [59], где обоснованы критерии оценки эффективности и безопасности нехирургических методов удаления локальных жировых отложений. Критерии рассмат-

риваются как инструмент стандартизации оценки результатов. Полученные данные позволяют повысить объективность сравнения методов.

Андрология представлена в работе [60], где описано улучшение способов диагностики и лечения секреторных нарушений фертильности у мужчин с олигоастенозооспермией. Олигоастенозооспермия рассматривается как состояние, характеризующееся снижением концентрации и подвижности сперматозоидов. Авторы предлагают усовершенствованные диагностические и лечебные подходы.

Управление безопасностью полётов представлено в исследовании [61], где представлено применение методологии контроллинга в управлении безопасностью полётов. Контроллинг рассматривается как интегрированная система информационно-аналитической поддержки управленческих решений. Полученные данные обосновывают применение методологии в авиационной сфере.

Управление персоналом представлено в работе [62], где описаны современные механизмы мотивации персонала инновационного предприятия. Мотивация рассматривается как ключевой фактор, определяющий эффективность инновационной деятельности. Авторы систематизируют механизмы и предлагают рекомендации по их применению.

Офтальмология представлена в исследовании [63], где исследованы полиморбидность и патогенетические основы формирования близорукости. Полиморбидность рассматривается как фактор, усугубляющий течение близорукости. Полученные данные расширяют представления о патогенезе и обосновывают комплексный подход к лечению.

Кардиология представлена в работе [64], где проведено изучение суточного профиля артериального давления, показателей ригидности сосудистой стенки и центрального аортального давления у больных гипертрофической кардиомиопатией. Суточное мониторирование рассматривается как метод, позволяющий оценить вариабельность показателей. Полученные данные имеют значение для стратификации риска.

Гепатология представлена в исследовании [65], где осуществлено исследование больных клинически различными формами хронического гепатита С на функциональную активность дендритных клеток. Дендритные клетки рассматриваются как ключевые регуляторы иммунного ответа. Полученные данные позволяют расширить представления о иммунопатогенезе.

Образование представлено в работе [66], где исследованы социальная составляющая концепции ESG и возможности её применения в государственных образовательных учреждениях. Концепция ESG рассматривается как интегральная модель устойчивого развития. Авторы анализируют возможности адаптации принципов ESG к образовательной сфере.

Неврология представлена в исследовании [67], где описано изучение состояния маркеров воспаления у больных с различными формами нарушения мозгового кровообращения. Маркеры воспаления рассматриваются как показатели, отражающие активность патологического процесса. Полученные дан-

ные имеют значение для дифференциальной диагностики.

Нейрореабилитация представлена в работе [68], где обосновано совершенствование нейрореабилитации больных с грубой степенью тяжести афазий. Афазия рассматривается как нарушение речи, требующее длительной и интенсивной реабилитации. Авторы предлагают усовершенствованные методы восстановления.

Дерматология представлена в исследовании [69], где доказано повышение эффективности лечения псориаза, ассоциированного с псориатическим артритом, путём включения в стандартный протокол экстракорпоральной фотохимиотерапии. Экстракорпоральная фотохимиотерапия рассматривается как метод, обеспечивающий иммуномодулирующий эффект. Полученные данные обосновывают включение данного метода в стандартный протокол.

Список использованных источников

1. Агаева М.И. Беременность и возможности прогнозирования активации латентной герпесвирусной инфекции// Журнал экспериментальной, клинической и профилактической медицины, №3(106), 2022. – С. 26-38
2. Алиева Ш.Р. Заболеваемость и смертность от “triple-negative” формы рака молочной железы на Апшеронском полуострове Азербайджанской Республики// Журнал экспериментальной, клинической и профилактической медицины, №1(104), 2022. – С. 4-10
3. Аршакян А.К. Прегравидарная подготовка пациенток с привычным невынашиванием беременности и хроническим эндометритом// Журнал экспериментальной, клинической и профилактической медицины, №4(107), 2022. – С. 4-20
4. Атласов И.В., Дубовицкая Т.В., Потудинский А.В. Оперативное управление вероятностными процессами в многостадийной специализированной системе// Экономика и менеджмент систем управления, №1(47), 2023. – С. 68-79
5. Атласов И.В., Лоренс С.А. Аналитические основы управления потоками системы оперативного оповещения с нестационарной структурой// Экономика и менеджмент систем управления, №2(48), 2023. – С. 59-67
6. Баймуканов А.М. Интервенционное лечение больных с фибрилляцией предсердий с учетом влияния восстановления и удержания синусового ритма на диастолическую функцию сердца// Журнал экспериментальной, клинической и профилактической медицины, №2(109), 2023. – С. 26-42
7. Баракина О.В. Улучшение результатов лечения пациенток репродуктивного возраста с хроническим эндометритом на основе разработанного и патогенетически обоснованного комплексного подхода с внутриматочным применением углеродного формованного сорбента, модифицированного поливинилпирролидоном// Журнал экспериментальной, клинической и профилактической медицины, №4(107), 2022. – С. 20-35
8. Бехбудова Л.Х. Изучение экспрессии генов рецепторов стероидных гормонов в мононуклеарных клетках периферической крови пациенток с миомой матки// Журнал экспериментальной, клинической и профилактической медицины, №1(108), 2023. – С. 4-20
9. Бикбаева Т.В. Прогностические критерии развития разных вариантов течения вэб-инфекционного мононуклеоза и эффективность иммунокорректирующей терапии больных// Журнал экспериментальной, клинической и профилактической медицины, №1(108), 2023. – С. 20-30
10. Бородкин А.В. Больные хронической обструктивной болезнью лёгких и хронической сердечной недостаточностью: особенности артериальной ригидности// Журнал экспериментальной, клинической и профилактической медицины, №3(110), 2023. – С. 3-20

11. Бывших Д.М., Жуков А.М., Перцев Ю.А. Особенности экономических рисков при серийных закупках техники радиоэлектронной борьбы// Экономика и менеджмент систем управления, №1(47), 2023. – С. 80-89
 12. Быкова А.П. Изменения толстой кишки при хроническом панкреатите и их клинико-прогностическое значение// Журнал экспериментальной, клинической и профилактической медицины, №4(111), 2023. – С. 4-20
 13. Велиева Г.В. Этиологические особенности дакриоцистита новорожденных// Журнал экспериментальной, клинической и профилактической медицины, №2(105), 2022. – С. 49-55
 14. Воронина Д.Д. Реабилитация пациентов после оперативного лечения грыж межпозвонковых дисков пояснично-крестцового отдела позвоночника – роль общей магнитотерапии// Журнал экспериментальной, клинической и профилактической медицины, №3(110), 2023. – С. 20-40
 15. Габриэль С.А., Хусаинова С.С., Слипченко А.С., Дынько В.Ю. Первые результаты радиочастотной абляции при лечении пищевода Барретта в краевой клинической больнице №2// Журнал экспериментальной, клинической и профилактической медицины, №3(106), 2022. – С. 39-45
 16. Галашокян К.М. Вакуум-терапия при хирургическом лечении эпителиального копчикового хода// Журнал экспериментальной, клинической и профилактической медицины, №3(106), 2022. – С. 45-61
 17. Гасым-заде И.Ш. Заболеваемость и смертность от рака легкого в Азербайджанской Республике// Журнал экспериментальной, клинической и профилактической медицины, №3(106), 2022. – С. 62-68
 18. Гафарова Л.А. Анализ влияния прогрессивности подоходного налогообложения на динамику экономических процессов// Экономика и менеджмент систем управления, №1(47), 2023. – С. 4-12
 19. Глаголева О.Н. Анемии, связанные с питанием: гигиеническое обоснование первичной профилактики// Журнал экспериментальной, клинической и профилактической медицины, №2(109), 2023. – С. 4-18
 20. Горелова Е.А. Течение, исходы, особенности ведения в острый и восстановительный периоды ротавирусной инфекции у детей с атопическим дерматитом// Журнал экспериментальной, клинической и профилактической медицины, №2(109), 2023. – С. 42-63
 21. Гребенникова Н.И., Сергеев М.Ю., Сергеева Т.И. Автоматизированная система рейтинговой оценки сервисных услуг фирмы// Экономика и менеджмент систем управления, №2(48), 2023. – С. 67-74
 22. Громов И.В. Анализ отдаленных результатов коррекции врожденной косолапости после использования подходов Виленского-Зацепина и Понсети у детей раннего возраста// Журнал экспериментальной, клинической и профилактической медицины, №4(107), 2022. – С. 35-48
 23. Елисеев А.В. Улучшение диагностики, непосредственных и отдаленных результатов хирургического лечения больных билиарным панкреатитом// Журнал экспериментальной, клинической и профилактической медицины, №4(107), 2022. – С. 49-72
 24. Ешиев А.М., Эшматов А.А., Сабиров С.А. Зависимость тактики лечения травматических повреждений челюстно-лицевой области от характера травмы// Журнал экспериментальной, клинической и профилактической медицины, №1(104), 2022. – С. 39-45
 25. Землянухин А.С., Мельников А.В., Моисеев С.И., Морозов В.П. Оценки показателей социально-экономического развития регионов на основе теории латентных переменных// Экономика и менеджмент систем управления, №1(47), 2023. – С. 12-21
 26. Ибишова Л.М. Ситуационный анализ заболеваемости раком яичников в Азербайджанской Республике// Журнал экспериментальной, клинической и профилактической медицины, №3(106), 2022. – С. 68-75
-

27. Иванков А.П., Селиверстов П.В. Оценка распространенности предикторов субхондрального стрессового перелома от недостаточности мышечков коленного сустава по данным магнитно-резонансной томографии// Журнал экспериментальной, клинической и профилактической медицины, №1(108), 2023. – С. 31-37
28. Кафтулина Ю.А., Смагина А.А. Современное состояние глобальных конфликтов: Восток – Запад, Север – Юг// Экономика и менеджмент систем управления, №1(47), 2023. – С. 21-28
29. Кафтулина Ю.А., Дубина А.Ш. Экономические и политические аспекты западных центров силы в управлении современной постбиополярной системой международных отношений// Экономика и менеджмент систем управления, №2(48), 2023. – С. 4-11
30. Кафтулина Ю.А., Смагина А.А. Международные тренды процесса цифровизации финансового рынка России// Экономика и менеджмент систем управления, №2(48), 2023. – С. 11-17
31. Ковалевская Е.А. Особенности ишемии миокарда у пациентов с гипертрофической кардиомиопатией// Журнал экспериментальной, клинической и профилактической медицины, №3(110), 2023. – С. 41-60
32. Ковалевская О.И. Клинико-лабораторные критерии острой цитомегаловирусной инфекции и результативность терапии у взрослых неиммунокомпрометированных больных// Журнал экспериментальной, клинической и профилактической медицины, №3(106), 2022. – С. 4-14
33. Козлов В.С., Дятлов В.В. Регулирование тарифов пассажирских перевозок в условиях современных вызовов// Экономика и менеджмент систем управления, №2(48), 2023. – С. 18-29
34. Кондратюк И.В. Патогенез эмоциональных нарушений при беременности с гипертензивными расстройствами// Журнал экспериментальной, клинической и профилактической медицины, №3(106), 2022. – С. 75-80
35. Королев С.В. Лечение деструктивного панкреатита с учетом энтерального зондового питания и селективной деконтаминации кишечника// Журнал экспериментальной, клинической и профилактической медицины, №1(104), 2022. – С. 46-58
36. Кравец О.Я. Особенности работы с упреждающими данными// Экономика и менеджмент систем управления, №2(48), 2023. – С. 74-83
37. Кудряшов С.И. Информативные лабораторные показатели прогнозирования трансформации острого гломерулонефрита в хронический на примере постинфекционного гломерулонефрита// Журнал экспериментальной, клинической и профилактической медицины, №4(111), 2023. – С. 20-36
38. Лазарев П.В. Роль анизоцитоза в сравнении с классическими факторами риска и маркерами воспаления в оценке прогноза заболевания при декомпенсации хронической сердечной недостаточности// Журнал экспериментальной, клинической и профилактической медицины, №1(104), 2022. – С. 59-75
39. Ларина Е.Б. Клинико-морфологическое обоснование применения антикоагулянтной терапии в профилактике плацентарной недостаточности// Журнал экспериментальной, клинической и профилактической медицины, №4(111), 2023. – С. 36-49
40. Лобанов О.С., Путькина Л.В., Минаков В.Ф. Математическая модель влияния инвестиций в когнитивные технологии на рост производительности труда// Экономика и менеджмент систем управления, №2(48), 2023. – С. 83-92
41. Мамедов М.К., Джавадзаде С.Н., Халафова Л.П. Об особенностях распространения и течения инфекций, вызванных вирусами гепатитов В и С среди онкологических больных, живущих в Азербайджане// Журнал экспериментальной, клинической и профилактической медицины, №2(109), 2023. – С. 18-26
42. Матвеев А.Г., Менухова Т.А. Момент возникновения потребности в товаре, как момент возникновения вероятности боя// Экономика и менеджмент систем управления, №1(47), 2023. – С. 28-37
-
-

43. Мельникова О.В., Ощепкова Н.Г. Экологический туризм как перспективное направление развития туристской отрасли Алтайского края// Экономика и менеджмент систем управления, №1(47), 2023. – С. 37-49
44. Мигунова М.И., Демина Л.Н., Меньшикова В.К., Могилевская Н.В. Обеспечение экономической безопасности предприятий в условиях ограничений в сфере внешнеэкономической деятельности// Экономика и менеджмент систем управления, №2(48), 2023. – С. 29-36
45. Мирза-заде Р.Ф. Эндоназальный электрофорез дексаметазона в комплексном лечении увеитов при болезни Бехтерева// Журнал экспериментальной, клинической и профилактической медицины, №4(107), 2022. – С. 72-79
46. Мирзоян А.Т. Улучшение результатов лечения больных с острым аппендицитом путем оптимизации техники лапароскопической аппендэктомии// Журнал экспериментальной, клинической и профилактической медицины, №4(111), 2023. – С. 49-63
47. Морозов Д.В. Исследование влияния наследственности на характер эпилепсии и эпилептоформных аномалий у детей// Журнал экспериментальной, клинической и профилактической медицины, №2(105), 2022. – С. 31-43
48. Мухаметова А.Д. Управление региональным развитием с учетом межрегиональных диспропорций// Экономика и менеджмент систем управления, №2(48), 2023. – С. 36-42
49. Норкулов Н.У., Шодиев А.Ш., Равшанов Д.М. Оценка эффективности применения ноотропов при лечении в остром периоде сотрясения головного мозга// Журнал экспериментальной, клинической и профилактической медицины, №2(105), 2022. – С. 43-48
50. Полякова М.А. Ранние стадии диабетической оптической нейропатии: патогенетические и клинко-диагностические аспекты// Журнал экспериментальной, клинической и профилактической медицины, №1(104), 2022. – С. 10-26
51. Прохоров В.В., Рожнов И.П. Оценка потенциала механизмов государственно-частного партнерства// Экономика и менеджмент систем управления, №1(47), 2023. – С. 49-60
52. Рогозянская М.И. Современные подходы к выполнению спленэктомии на этапе хирургического лечения рака желудка 37-42
53. Рогозянская М.И., Редькин А.Н. Сравнительный анализ и пути улучшения результатов гастрэктомии со спленэктомией и без спленэктомии у больных раком желудка// Журнал экспериментальной, клинической и профилактической медицины, №2(109), 2023. – С. 63-69
54. Рябенко Э.Б., Рябенко Ю.Н. Анализ заболеваемости лихорадкой Западного Нила в Российской Федерации и мире в 2020 году// Журнал экспериментальной, клинической и профилактической медицины, №1(104), 2022. – С. 76-80
55. Сапожникова Ю.А. Оценка морфологии кожи у больных с хроническими дерматозами на основе ультразвукового сканирования// Журнал экспериментальной, клинической и профилактической медицины, №2(105), 2022. – С. 4-18
56. Сидоренко К.И., Козлов В.С. Методические аспекты организации деятельности в сфере оказания транспортных услуг// Экономика и менеджмент систем управления, №2(48), 2023. – С. 42-52
57. Сотников И.А. Проблема идентификации цитомегаловирусной инфекции, а также ее влияния на формирование соматической патологии у детей// Журнал экспериментальной, клинической и профилактической медицины, №3(106), 2022. – С. 14-25
58. Суркичин С.И. Обоснование применения и подходы к терапии локальных жировых отложений с помощью аппаратных методов – криолиполиза и радиоволнового воздействия// Журнал экспериментальной, клинической и профилактической медицины, №1(108), 2023. – С. 42-58
59. Суркичин С.И. Нехирургические методы удаления локальных жировых отложений: критерии оценки эффективности и безопасности// Журнал экспериментальной, кли-
-

нической и профилактической медицины, №2(109), 2023. – С. 69-80

60. Фесенко С.В. Улучшение способов диагностики и лечения секреторных нарушений фертильности у мужчин с олигоастенозооспермией// Журнал экспериментальной, клинической и профилактической медицины, №2(105), 2022. – С. 55-67

61. Филько С.В., Кузнецов Е.В. Применение методологии контроллинга в управлении безопасностью полетов// Экономика и менеджмент систем управления, №2(48), 2023. – С. 94-100

62. Финошкин О.В. Современные механизмы мотивации персонала инновационного предприятия// Экономика и менеджмент систем управления, №1(47), 2023. – С. 60-67

63. Хамнагдаева Н.В. Полиморбидность и патогенетические основы формирования близорукости// Журнал экспериментальной, клинической и профилактической медицины, №3(110), 2023. – С. 61-78

64. Хашиева Ф.М. Изучение суточного профиля артериального давления, показателей ригидности сосудистой стенки, центрального аортального давления у больных гипертрофической кардиомиопатией// Журнал экспериментальной, клинической и профилактической медицины, №1(108), 2023. – С. 59-68

65. Цветков В.В. Исследование больных клинически различными формами хронического гепатита С на функциональную активность дендритных клеток// Журнал экспериментальной, клинической и профилактической медицины, №1(104), 2022. – С. 26-38

66. Четин А.М. Социальная составляющая концепции ESG и возможности ее применения в государственных образовательных учреждениях// Экономика и менеджмент систем управления, №2(48), 2023. – С. 52-58

67. Чипова Д.Т. Изучение состояния маркеров воспаления у больных с различными формами нарушения мозгового кровообращения// Журнал экспериментальной, клинической и профилактической медицины, №2(105), 2022. – С. 18-30

68. Щербакова М.М. Совершенствование нейрореабилитации больных с грубой степенью тяжести афазий// Журнал экспериментальной, клинической и профилактической медицины, №4(111), 2023. – С. 63-80

69. Якубовская Е.С. Повышение эффективности лечения псориаза, ассоциированного с псориатическим артритом путем включения в стандартный протокол экстракорпоральной фотохимиотерапии// Журнал экспериментальной, клинической и профилактической медицины, №2(105), 2022. – С. 67-80

Немцов Л.Б., Немцов М.Л.

СКВОЗНАЯ НАБЛЮДАЕМОСТЬ ERP-СИСТЕМ

**Санкт-Петербургский государственный электротехнический
университет «ЛЭТИ» им. В.И. Ульянова (Ленина)**

Введение

Современные ERP-системы являются основой цифровой инфраструктуры предприятий и обеспечивают выполнение критически важных бизнес-процессов. Рост сложности архитектуры ERP, широкое использование распределенных вычислений и повышение требований к непрерывности функционирования делают задачу контроля состояния таких систем одной из ключевых задач эксплуатации. Дополнительную актуальность данной проблеме придает включение ERP-систем в перечень объектов критической информационной инфраструктуры Российской Федерации, что существенно повышает требования к обеспечению их надежности и наблюдаемости.

В работе рассматривается текущее состояние наблюдаемости, монито-

ринга и анализа ERP-систем, и предлагается концепция построения наблюдаемости, формируется понятийный аппарат и определяются основные принципы построения моделей наблюдаемости ERP-ландшафта.

Предлагаемая концепция предполагает перенос объекта анализа с источников телеметрии на сущности наблюдаемости. Телеметрические данные рассматриваются не как самостоятельный объект анализа, а как средство получения информации о состоянии сущностей ERP-ландшафта.

1. Обзор текущего состояния наблюдаемости

Анализ научной и профессиональной литературы показывает, что современная проблематика наблюдаемости (Observability) информационных систем сформировалась преимущественно в контексте распределенных программных систем, микросервисных архитектур, DevOps/SRE-практик и AIOps. В работах С. Sridharan, С. Majors, L. Fong-Jones и G. Miranda наблюдаемость рассматривается как концептуальное развитие традиционного мониторинга, ориентированное на исследование внутреннего состояния сложной программной системы по внешним проявлениям ее функционирования, на основе внешне доступных данных телеметрии: метрик, логов и трассировок [1, 2]. В книге Google Site Reliability Engineering мониторинг распределенных систем рассматривается как основа обеспечения надежности и доступности крупных программных сервисов [3]. Близкая позиция представлена в работе S. Niedermaier и соавторов, где на основе интервью с промышленными специалистами показано, что рост сложности распределенных систем требует не только технических инструментов мониторинга, но и организационно оформленной практики наблюдаемости [4].

Формирование современных подходов к наблюдаемости тесно связано с развитием распределенной трассировки. В системах X-Trace, Dapper, Pivot Tracing и Sapору были предложены методы восстановления связей между операциями, выполняемыми различными компонентами распределенной системы, и формирования сквозного представления о пути обработки запроса [5-8]. Дальнейшие исследования развивают трассировку и методы автоматического извлечения из трассировок графов зависимостей, метрик и последовательностей выполнения операций [9, 10].

Существенное влияние на развитие области оказала стандартизация телеметрии. Спецификация OpenTelemetry определяет единый подход к сбору, представлению и передаче данных наблюдаемости, включая трассировки, метрики, логи и контекст выполнения [11]. Однако OpenTelemetry решает преимущественно задачу унификации инструментирования и обмена телеметрическими данными, не определяя предметную модель объектов наблюдаемости и их взаимосвязей в корпоративной информационной системе.

Отдельное направление исследований связано с машинным обучением и анализом больших данных - AIOps и автоматизированным анализом инцидентов. В обзоре P. Notaro, J. Cardoso и M. Gerndt систематизированы методы AIOps для управления отказами, включая обнаружение аномалий, корреляцию событий, локализацию неисправностей и анализ первопричин [12]. Со-

временные работы по поиску причин возникновения инцидентов (Root Cause Analysis) развивают методы интерпретируемого анализа событий и мультимодальной телеметрии, однако объектом анализа в них, как правило, выступают микросервисы, инфраструктурные компоненты или программные сервисы, а не ERP-система как совокупность взаимосвязанных бизнес-, прикладных, платформенных и инфраструктурных сущностей [13-16].

В смежной области Business Process Management (BPM) и Business Activity Monitoring (BAM) основное внимание уделяется моделированию, исполнению и контролю бизнес-процессов. W.M.P. van der Aalst рассматривает BPM как совокупность методов, техник и инструментов для проектирования, выполнения, анализа и улучшения бизнес-процессов [17]. Данное направление важно для исследования ERP-систем, поскольку ERP-система непосредственно поддерживает выполнение сквозных бизнес-процессов предприятия. Вместе с тем BPM и BAM не формируют единой модели технической и бизнес-наблюдаемости ERP-системы, объединяющей данные бизнес-процессов, прикладных объектов, платформы, инфраструктуры и телеметрии.

Близким направлением является Аналитика процессов (Process Mining), в рамках которого журналы событий используются для обнаружения, контроля соответствия и анализа фактического выполнения бизнес-процессов [18, 19]. Однако здесь объектом выступает процесс и последовательность его событий, тогда как состояние программной платформы, базы данных, вычислительной и сетевой инфраструктуры обычно остается за пределами модели анализа.

Исследования ERP-систем в основном сосредоточены на вопросах их эволюции, внедрения, интеграции, организационного влияния и качества управления предприятием [20, 21]. Так, S. Katuu рассматривает развитие ERP-систем от MRP и MRP II к современным интегрированным корпоративным системам [22]. Исследования также показывают, что результативность эксплуатации ERP определяется соответствием между функциональными возможностями корпоративной системы и организационной структурой предприятия [23]. Вместе с тем данные работы не рассматривают ERP-ландшафт как объект сквозной технической и бизнес-наблюдаемости.

В отечественной литературе ERP-системы также чаще всего анализируются с позиции структуры, функционального состава, внедрения и организационной эффективности, например, в широко цитируемой работе А.А. Григорьева и В.А. Титова [24].

Ранее автором были исследованы методологические аспекты внедрения 1С на производственном предприятии [25]. Однако как зарубежные, так и отечественные исследования ERP-систем, как правило, не рассматривают наблюдаемость ERP как самостоятельный объект научного анализа.

Особое значение данная проблема приобретает в российском контексте. Распоряжением Правительства Российской Федерации от 26 февраля 2026 г. № 360-р утвержден перечень типовых отраслевых объектов критической информационной инфраструктуры Российской Федерации [26]. Согласно по-

становлению, ERP-системы включены в объекты критической информационной инфраструктуры. В условиях использования ERP-систем на промышленных предприятиях и в критически значимых отраслях задача сквозного контроля состояния таких систем выходит за рамки традиционного администрирования и становится частью обеспечения устойчивости корпоративных информационных систем и экономической безопасности предприятия в целом.

Анализ отечественных публикаций позволяет сделать вывод, что в российской научной печати тематика наблюдаемости ERP-систем как самостоятельного направления практически не представлена. Выявленные работы преимущественно используют категории мониторинга, эксплуатации, сопровождения, внедрения, цифровизации и управления корпоративными информационными системами. При этом концепция сквозной наблюдаемости ERP-систем, основанная на представлении ERP как совокупности взаимосвязанных сущностей наблюдаемости, в доступной российской научной литературе авторами не выявлена.

Таким образом, существующие подходы развивают отдельные аспекты мониторинга, наблюдаемости, телеметрии, AIOps, BPM и ERP-управления, однако не предлагают единой концептуальной модели, в которой объектами наблюдаемости выступают взаимосвязанные сущности ERP-системы. При этом во всех рассмотренных подходах первичным объектом анализа выступают программные компоненты, сервисы, процессы или источники телеметрии, тогда как единая модель предметных сущностей ERP-системы как объектов наблюдаемости отсутствует.

2. Проблемы существующих подходов к наблюдаемости

Проведенный анализ публикаций показывает, что существующие подходы к мониторингу и наблюдаемости решают широкий спектр задач эксплуатации сложных информационных систем, однако их применение к крупным ERP-системам имеет ряд принципиальных ограничений.

Традиционные системы мониторинга ориентированы преимущественно на контроль состояния отдельных компонентов информационной инфраструктуры: серверов, сетевого оборудования, баз данных, виртуальных машин и операционных систем. Такой подход обеспечивает своевременное обнаружение отказов оборудования и программной платформы, однако практически не отражает влияние технических инцидентов на функционирование прикладных компонентов ERP-системы и выполнение бизнес-процессов.

Концепция *наблюдаемости* (Observability) существенно расширила возможности мониторинга за счет использования телеметрических данных, включающих метрики, журналы событий и распределенные трассировки. Это позволило анализировать внутреннее состояние распределенных приложений, выявлять сложные причинно-следственные связи между программными сервисами и повысить эффективность диагностики неисправностей. Вместе с тем объектом наблюдаемости в большинстве существующих решений остаются программные компоненты, микросервисы, контейнеры, экземпляры приложений и другие технические элементы вычислительной инфраструкту-

ры.

Развитие методов AIOps позволило автоматизировать обработку больших объемов телеметрических данных, корреляцию событий, выявление аномалий и поиск первопричин инцидентов. Однако эффективность подобных методов во многом определяется полнотой контекста наблюдаемых объектов. При отсутствии формализованной модели предметной области ERP алгоритмы машинного обучения вынуждены работать преимущественно с технической телеметрией, что затрудняет интерпретацию результатов с точки зрения влияния инцидентов на прикладную функциональность и бизнес-процессы предприятия.

Смежные направления исследований, связанные с управлением и мониторингом бизнес-процессов, рассматривают выполнение операций предприятия как основной объект анализа. Однако данные подходы, как правило, абстрагируются от состояния программной платформы, инфраструктуры и прикладных компонентов ERP-системы. В результате технические и бизнес-аспекты функционирования информационной системы анализируются независимо друг от друга, что существенно усложняет локализацию причин возникновения инцидентов и оценку их влияния на деятельность предприятия.

Особую сложность представляет многоуровневая архитектура современных ERP-систем. На практике выполнение одной бизнес-операции одновременно зависит от состояния бизнес-процессов, прикладной конфигурации ERP, платформы исполнения, системы управления базами данных, вычислительной инфраструктуры и сетевого взаимодействия. Несмотря на наличие средств мониторинга каждого из перечисленных уровней, существующие решения, как правило, рассматривают их как самостоятельные объекты наблюдения, не формируя единой модели взаимосвязей между ними.

Следствием такого подхода является фрагментарность представления информации о состоянии ERP-системы. Метрики, журналы регистрации, трассировки, события платформы, данные операционных систем и сведения о выполнении бизнес-процессов хранятся в различных информационных источниках и анализируются разрозненно и, часто, разными специалистами. Корреляция этих данных чаще всего осуществляется на основе эвристических правил или экспертных знаний специалистов по эксплуатации, что затрудняет автоматизацию анализа и снижает эффективность диагностики сложных инцидентов.

Таким образом, основным ограничением существующих подходов является отсутствие единой концептуальной модели, позволяющей рассматривать ERP-систему как совокупность взаимосвязанных объектов наблюдаемости независимо от их уровня в архитектуре системы. В настоящее время отсутствует единая абстракция, которая позволяла бы одинаковым образом описывать бизнес-процессы, прикладные объекты ERP, компоненты платформы, элементы инфраструктуры и связанные с ними данные телеметрии.

Для устранения указанного ограничения предлагается использовать сквозную наблюдаемость ERP-систем на основе сущностей. В отличие от

существующих подходов, базовым объектом анализа выступает сущность наблюдаемости (Observability Entity) - идентифицируемый объект ERP-системы, обладающий состоянием, наблюдаемыми характеристиками и множеством связей с другими сущностями. Такой подход обеспечивает единое представление объектов различных уровней архитектуры ERP-системы и создает основу для последующей корреляции технических и бизнес-событий в рамках общей модели сквозной наблюдаемости.

В разрабатываемой концепции объектом наблюдаемости становится не источник телеметрии и не программный компонент, а сущность наблюдаемости, представляющая любой идентифицируемый объект ERP-ландшафта независимо от уровня его расположения в архитектуре.

Под ERP-ландшафтом будем понимать совокупность взаимосвязанных аппаратных, инфраструктурных, платформенных, прикладных и бизнес-компонентов, размещенных на пяти уровнях абстракций функционирования ERP-системы.

3. Разработка концепции сквозной наблюдаемости

Основным положением разрабатываемой концепции является перенос объекта наблюдаемости с отдельных технических компонентов и источников телеметрии на формализованные сущности наблюдаемости. ERP-система при таком подходе представляется как совокупность взаимосвязанных сущностей, каждая из которых может обладать собственным состоянием, наблюдаемыми характеристиками и отношениями с другими сущностями.

Сущностью наблюдаемости может выступать любой значимый объект ERP-ландшафта независимо от его физической или логической природы и уровня расположения в архитектуре. Такими объектами могут являться бизнес-процесс, бизнес-операция, прикладной объект, документ, справочник, регистр, пользовательский сеанс, рабочий процесс платформы, сервер кластера приложений, экземпляр системы управления базами данных, виртуальная машина, физический сервер или другой инфраструктурный компонент.

Каждая сущность наблюдаемости может характеризоваться совокупностью следующих элементов:

- уникальным идентификатором, обеспечивающим однозначное различение сущности в пределах наблюдаемой ERP-системы;
 - типом, определяющим принадлежность сущности к определенному классу объектов;
 - состоянием, отражающим текущее или историческое положение сущности;
 - набором наблюдаемых характеристик, позволяющих оценивать ее состояние;
 - источниками данных наблюдаемости, из которых могут быть получены сведения о сущности;
 - контекстом, содержащим дополнительную информацию, необходимую для интерпретации ее состояния;
 - множеством отношений с другими сущностями наблюдаемости.
-
-

Таким образом, метрика характеризует количественно измеряемое свойство сущности; запись журнала отражает произошедшее с сущностью событие или изменение ее состояния; трассировка характеризует участие сущности в выполнении распределенной операции; событие фиксирует значимый факт, связанный с ее функционированием. Один объект ERP-системы может иметь несколько источников данных наблюдаемости, а один источник данных при соответствующей идентификации и корреляции может содержать информацию о нескольких взаимосвязанных сущностях.

Данный подход позволит перейти от анализа изолированных телеметрических сигналов к формированию целостного представления о состоянии ERP-системы. Например, снижение производительности конкретной бизнес-операции может быть связано с прикладным объектом ERP-системы, пользовательским сеансом, рабочим процессом платформы, сервером приложений, экземпляром СУБД и инфраструктурными компонентами. Представление перечисленных объектов как взаимосвязанных сущностей наблюдаемости создает основу для построения сквозной цепочки зависимостей между бизнес-состоянием и техническим состоянием системы.

Таким образом каждая значимая сущность системы должна иметь возможность быть однозначно идентифицирована, наблюдаема посредством доступных источников данных и связана с другими сущностями независимо от уровня архитектуры, на котором она расположена.

Реализация данного принципа требует сохранения контекста наблюдаемости при переходе между различными уровнями ERP-системы. Контекст наблюдаемости обеспечивает возможность установить, например, связь между конкретной бизнес-операцией, сформированным в результате ее выполнения документом ERP-системы, пользовательским сеансом, рабочим процессом платформы, обращением к базе данных и соответствующими инфраструктурными компонентами. Благодаря этому данные различных источников перестают рассматриваться как независимые сигналы и могут быть объединены вокруг конкретных сущностей и отношений между ними.

Таким образом, разрабатываемая концепция должна основываться на четырех ключевых положениях. Во-первых, первичным объектом наблюдаемости является сущность, а не отдельный источник телеметрических данных. Во-вторых, сущности различных уровней ERP-системы описываются посредством единого понятийного аппарата. В-третьих, данные наблюдаемости связываются с конкретными сущностями и характеризуют их состояние. В-четвертых, отношения между сущностями обеспечивают сквозной переход от бизнес-процессов и прикладных объектов к платформенным и инфраструктурным компонентам.

Основополагающие принципы создают основу для представления ERP в виде системы взаимосвязанных сущностей наблюдаемости. Структурированное описание таких сущностей опирается на пятиуровневую модель, включающую бизнес-, прикладной, технологический, инфраструктурный и физический уровни (рис. 1).



Рис. 1. Пятиуровневый ERP-ландшафт

Физический уровень объединяет аппаратные компоненты, обеспечивающие функционирование информационной системы: вычислительные серверы, процессоры, оперативную память, системы хранения данных, сетевое оборудование и каналы связи.

Инфраструктурный уровень включает программные компоненты, обеспечивающие выполнение прикладных сервисов: операционные системы, средства виртуализации, контейнерные платформы, серверы приложений, системы управления базами данных, службы каталогов, сетевые сервисы и промежуточное программное обеспечение.

Технологический уровень отражает особенности платформы исполнения ERP-системы и механизмы обработки данных.

Прикладной уровень содержит объекты прикладной конфигурации ERP-системы: документы, справочники, регистры, отчеты, обработки, роли пользователей, прикладные сервисы и интеграционные механизмы.

Бизнес-уровень представляет наиболее высокий уровень абстракции и включает бизнес-процессы, бизнес-операции, показатели эффективности (KPI), соглашения об уровне обслуживания (SLA), аналитические показатели и другие сущности, характеризующие деятельность предприятия и результаты функционирования ERP-системы.

Предлагаемое разделение носит концептуальный характер и не зависит от конкретной программной платформы.

Заключение

В статье рассмотрена основа концепции сквозной наблюдаемости ERP-систем на основе сущностей наблюдаемости, ориентированная на преодоление ограничений существующих подходов к мониторингу и наблюдаемости корпоративных информационных систем. Проведенный анализ отечественных и зарубежных исследований показал, что современные методы мониторинга, observability, AIOps и управления бизнес-процессами решают отдельные задачи эксплуатации распределенных программных систем, однако не формируют единой концептуальной модели наблюдаемости ERP-систем.

Новизна предлагаемого подхода заключается во введении понятия сущности наблюдаемости (Observability Entity) как универсального объекта описания ERP-ландшафта. В отличие от существующих подходов, первичным

объектом анализа выступают не программные компоненты и не источники телеметрии, а формализованные сущности наблюдаемости, характеризующие объекты различных уровней архитектуры ERP-системы. Телеметрические данные при этом рассматриваются как средство получения информации о состоянии сущностей, а не как самостоятельный объект анализа.

Концепция рассматривает ERP-систему не как изолированное программное приложение, а как ERP-ландшафт, включающий взаимосвязанные физические, инфраструктурные, технологические, прикладные и бизнес-компоненты. Представление ERP-ландшафта в виде совокупности сущностей наблюдаемости создает основу для формирования единого пространства наблюдаемости и установления причинно-следственных связей между событиями различных уровней архитектуры.

Концепция определяет понятийный аппарат сквозной наблюдаемости ERP-систем. Ее дальнейшее развитие связано с построением типизированной многослойной модели ERP-ландшафта, в которой сущности наблюдаемости и отношения между ними получают строгое формальное описание.

Дальнейшее развитие исследования предполагает разработку *типизированной многослойной модели сквозной наблюдаемости* ERP-систем и экспериментальную верификацию модели на высоконагруженных промышленных ERP-системах.

Список использованных источников

1. Sridharan C. Distributed Systems Observability: A Guide to Building Robust Systems. - Sebastopol, CA: O'Reilly Media, 2018.
 2. Majors C., Fong-Jones L., Miranda G. Observability Engineering: Achieving Production Excellence. - Sebastopol, CA: O'Reilly Media, 2022.
 3. Beyer B., Jones C., Petoff J., Murphy N.R. Site Reliability Engineering: How Google Runs Production Systems. - Sebastopol, CA: O'Reilly Media, 2016.
 4. Niedermaier S., Koetter F., Freymann A., Wagner S. On Observability and Monitoring of Distributed Systems: An Industry Interview Study// Lecture Notes in Business Information Processing. 2019. P. 36-52. DOI: 10.1007/978-3-030-33702-5_3.
 5. Dapper, a large-scale distributed systems tracing infrastructure/ B.H. Sigelman et al. - 2010. - <https://research.google/pubs/dapper-a-large-scale-distributed-systems-tracing-infrastructure>.
 6. X-Trace: a pervasive network tracing framework/ R. Fonseca et al.// Proc. of the 4th USENIX Symp. on Networked Systems Design and Implementation (NSDI'07). - Berkeley: USENIX Association, 2007. - P. 271-284.
 7. Pivot tracing: dynamic causal monitoring for distributed systems/ J. Mace, R. Roelke, R. Fonseca// Proc. of the 25th ACM Symp. on Operating Systems Principles (SOSP'15). - New York: ACM, 2015. - P. 378-393.
 8. Canopy: an end-to-end performance tracing and analysis system/ J. Kaldor et al.// Proc. of the 26th ACM Symp. on Operating Systems Principles (SOSP'17). - New York: ACM, 2017. - P. 34-50.
 9. Principled workflow-centric tracing of distributed systems/ R.R. Sambasivan et al.// Proc. of the 2016 ACM Symp. on Cloud Computing (SoCC'16). - New York: ACM, 2016. - P. 401-414.
 10. Automated analysis of distributed tracing: challenges and research directions/ A. Bento et al.// Journal of Grid Computing. - 2021. - Vol. 19. - Article 9. - DOI 10.1007/s10723-021-09551-5.
-

11. <https://opentelemetry.io/docs/specs/otel/>.
12. A survey of AIOps methods for failure management/ P. Notaro, J. Cardoso, M. Gerndt// ACM Transactions on Intelligent Systems and Technology. - 2021. - Vol. 12, no. 6. - Article 81. - DOI 10.1145/3483424.
13. Latent error prediction and fault localization for microservice applications by learning from system trace logs/ X. Zhou et al.// Proc. of the 27th ACM Joint European Software Engineering Conf. and Symp. on the Foundations of Software Engineering (ESEC/FSE'19). - New York: ACM, 2019. - P. 683-694. -
14. Anomaly detection from system tracing data using multimodal deep learning/ S. Nedelkoski, J. Cardoso, O. Kao// 2019 IEEE 12th Int. Conf. on Cloud Computing (CLOUD). - Piscataway: IEEE, 2019. - P. 179-186.
15. DeepLog: anomaly detection and diagnosis from system logs through deep learning/ M. Du, F. Li, G. Zheng, V. Srikumar// Proc. of the 2017 ACM SIGSAC Conf. on Computer and Communications Security (CCS'17). - New York: ACM, 2017. - P. 1285-1298.
16. Microservices monitoring with event logs and black box execution tracing/ M. Cinque, R. Della Corte, A. Pecchia// IEEE Transactions on Services Computing. - 2022. - Vol. 15, no. 1. - P. 294-307.
17. van der Aalst W.M.P. Business Process Management: A Comprehensive Survey// ISRN Software Engineering. 2013. Vol. 2013. Article ID 507984. DOI: 10.1155/2013/507984.
18. van der Aalst W.M.P. Process mining: overview and opportunities// ACM Transactions on Management Information Systems. - 2012. - Vol. 3, no. 2. - Article 7. - P. 1-17.
19. Process mining manifesto/ W.M.P. van der Aalst et al.// Lecture Notes in Business Information Processing, vol. 99, 2012. - P. 169-194.
20. What is ERP? / H. Klaus, M. Rosemann, G. G. Gable// Information Systems Frontiers. - 2000. - Vol. 2, no. 2. - P. 141-162.
21. Enterprise resource planning (ERP) - a brief history/ F.R. Jacobs, F.C. Weston// Journal of Operations Management. - 2007. - Vol. 25, no. 2. - P. 357-363.
22. Katuu S. Enterprise Resource Planning: Past, Present, and Future// New Review of Information Networking. 2020. DOI: 10.1080/13614576.2020.1742770.
23. Understanding organization-enterprise system fit: a path to theorizing the information technology artifact/ D.M. Strong, O. Volkoff// MIS Quarterly. - 2010. - Vol. 34, no. 4. - P. 731-756.
24. Григорьев А.А., Титов В.А. Характеристика, структура, организация систем управления ERP, ERP II и ERP III// Фундаментальные исследования. 2017. № 2. С. 48-51.
25. Немцов М.Л., Немцов Л.Б. Методология внедрения 1С:ERP на производственном предприятии// Системы управления и информационные технологии. 2024. № 4(98). С. 48-51.
26. Об утверждении перечня типовых отраслевых объектов критической информационной инфраструктуры Российской Федерации: распоряжение Правительства Российской Федерации от 26.02.2026 № 360-р// Собрание законодательства Российской Федерации. - 2026. - № 9.- Ст. 1048.

2. МОДЕЛИРОВАНИЕ И АНАЛИЗ СЛОЖНЫХ СИСТЕМ

Мадетбаева Н.Б.

ГИБРИДНАЯ МОДЕЛЬ БАЙЕСА И ГРАДИЕНТНОГО БУСТИНГА НА ОСНОВЕ ОНТОЛОГИИ УЗБЕКСКИХ ФРАЗЕОЛОГИЗМОВ

Научно-исследовательский институт развития цифровых технологий и искусственного интеллекта, Ташкент, Узбекистан.

Введение

Фразеологизмы, то есть устойчивые сочетания, значение которых не выводится из простой суммы значений входящих в них слов, представляют собой одну из старейших и наиболее трудных проблем автоматической обработки любого языка. В научной литературе особо подчёркивалось, что многословные выражения являются серьёзным препятствием для обработки естественного языка [1], а за последние двадцать лет это направление превратилось в самостоятельную область исследований [2]. В узбекском языке острота проблемы ещё возрастает. Во-первых, узбекский язык является агглютинативным, поэтому слова в составе фразеологизма за счёт цепочек аффиксов принимают бесчисленное множество форм; например, в предложении «boshini yerdan uzmađi» глагол «uzmoq» употреблён с аффиксами отрицания и времени. Эта особенность узбекского языка занимает центральное место и в средствах морфологического анализа [3, 4]. Кроме того, для систем машинного перевода и обработки текстов предложены эффективные алгоритмические стратегии морфологического анализа сложных узбекских слов и выделения их основ (стемминга) [5].

Во-вторых, компоненты фразеологизма не всегда образуют единый блок - между ними очень часто вставляются другие слова. Например, фразеологизм «zavqi toshib ketdi» встречается в тексте и в виде «zavqi juda toshib ketdi». В-третьих, ресурсы узбекского языка ограничены, то есть крупных размеченных данных практически нет, поэтому чисто статистические или чисто нейронные методы не имеют достаточного обеспечения. Тем не менее в последние годы активизируются прикладные исследования по созданию крупных размеченных корпусов и нейросетевых моделей для автоматической обработки текстов на узбекском языке [6].

Среди статистических подходов классическим инструментом считается критерий точечной взаимной информации (PMI), измеряющий связь между словами [7]. Однако PMI и подобные ему традиционные статистические критерии имеют ряд серьёзных недостатков при извлечении фразеологизмов. Опираясь только на частоту совместной встречаемости и не улавливая семантический сдвиг, они не отличают фразеологизмы от обычных коллокаций (например, «choy ichmoq»); из-за морфологической гибкости узбекского языка и вставки посторонних слов между компонентами показатели PMI резко падают; наконец, PMI является одномерным показателем и не способен комплексно моделировать нелинейные признаки - расстояние между словами, порядок, частоту и грамматические ограничения.

Для устранения этих недостатков и объединения различных статистических и структурных признаков в единое сильное решающее правило в качестве обучаемой модели выбран градиентный бустинг. Градиентный бустинг - метод, при котором множество небольших решающих деревьев обучаются последовательно, каждое из которых исправляет ошибки предыдущих, и из их суммы формируется сильная модель [8]. Его вероятностный вариант NGBoost предсказывает не одно число, а целое распределение вероятностей [9], что естественным образом связывается с байесовским подходом. В байесовском подходе знания, имеющиеся до анализа данных, задаются априорным распределением, которое после учёта данных переходит в апостериорное.

Поскольку в условиях ограниченных ресурсов чисто статистические показатели не отражают морфологическую изменчивость и разрывность фразеологизмов, при формировании байесовского априорного распределения используется база экспертных знаний, объединяющая лингвистические правила и семантические связи между компонентами. Под онтологией понимается формальное, машиночитаемое описание понятий предметной области и отношений между ними [10]. Онтология используется как источник априорного знания — кандидат, известный онтологии, получает высокое начальное доверие. Однако сама по себе она лишь хранилище знаний и не находит фразеологизм в тексте, поэтому основу исследования составляет объединение обучающейся на данных статистической модели и экспертного знания в одной системе. Для переноса графа знаний в непрерывное векторное пространство применён метод TransE. TransE стремится к тому, чтобы в каждом триплете сумма вектора головного узла и вектора отношения была близка к вектору хвостового узла [11].

Цель статьи состоит в математическом описании гибридной модели и её экспериментальной проверке на материале узбекского языка. Для этого модели А (гибрид байесовского подхода и градиентного бустинга, опирающийся только на статистику текста) и В (та же архитектура с онтологией узбекских фразеологизмов в качестве третьего компонента) сопоставлены в совершенно одинаковых условиях; поскольку единственным различием является онтология, любое превосходство модели В интерпретируется как её вклад.

Методология

Данные. В исследовании использованы три типа ресурсов:

Первый - объединённый текстовый корпус, собранный из восьми художественных произведений (два романа А. Кадыри, произведения Чулпана, Айбека, два произведения У. Хашимова, произведения Г. Гуляма и А. Исмаила) и выборки открытого узбекского корпуса; корпус содержит 154 502 предложения и более 1,07 млн лемма-токенов. Все предложения лемматизированы с помощью средства UzMorphAnalyser [3]. Этот корпус служит только для вычисления статистических показателей (частоты, PMI и других).

Второй - золотой стандарт, то есть размеченный экспертом набор из 860 предложений. В 399 из них имеется хотя бы один фразеологизм (положительный класс), в 461 предложении фразеологизмов нет (отрицательный класс); в положительных предложениях указана также текстовая форма фра-

зеологизма.

Третий - онтология узбекских фразеологизмов. Она представлена в формате OWL и состоит из 28 310 триплетов; включает 2 098 фразеологических единиц и 1 104 узла-леммы. Для каждого фразеологизма зафиксированы шаблон компонентов (componentPattern - последовательность лемм), опорная лемма (hasAnchorLemma - смысловое ядро фразеологизма), семантическая и структурная категории, а для лемм - допустимые группы аффиксов (allowedAffixGroup - допустимые типы аффиксов леммы).

Формирование кандидатов и статистические признаки. Каждое предложение золотого стандарта сначала разбивается на токены, после чего простым стеммером на основе правил отсекаются словоизменяющие и формообразующие аффиксы и выделяются основы. Использование математических моделей и алгоритмических средств играет важную роль в грамматически корректной обработке словоформ узбекских текстов и их разделении на основу и аффиксы [12]. Далее каждая упорядоченная пара слов (w_i, w_j) внутри предложения рассматривается как кандидат, если расстояние между ними не превышает трёх слов. Каждому кандидату соответствует истинная метка $y_c \in \{0,1\}$. Для $y_c = 1$ обе основы должны входить в состав фразеологизма, размеченного экспертом. Таким образом, из 860 предложений получено в общей сложности 21 084 кандидата, из которых 623 (3,0%) являются парами компонентов настоящих фразеологизмов.

Для каждого кандидата строится шестимерный вектор статистических признаков:

$$\mathbf{x}_c^{stat} = [\text{PMI}, \text{Freq}, \text{Skip}, \text{Dist}, \text{T}, \text{Order}]. \quad (1)$$

Здесь $\mathbf{x}_c^{stat}$ - шестимерный вектор статистических признаков пары слов кандидата c ;

PMI - точечная взаимная информация (Pointwise Mutual Information), измеряющая связь между словами;

Freq - частота совместной встречаемости слов кандидата в текстовом корпусе;

Skip - сумма встреч с обратным взвешиванием по расстоянию при вставке посторонних слов между компонентами;

Dist - фактическое расстояние между словами кандидата в контексте текста (число слов);

T - показатель T -теста, оценивающий неслучайность и статистическую значимость встречаемости пары;

Order - признак, отражающий порядок следования слов в тексте и наличие инверсии.

Точечная взаимная информация PMI, измеряющая связь между основами кандидата, вычисляется следующим образом:

$$\text{PMI}(a, b) = \log_2 \frac{P(a, b)}{P(a)P(b)}. \quad (2)$$

Здесь $P(a, b)$ - вероятность совместного появления основ a и b в контекстном окне из четырёх слов;

$P(a)$, $P(b)$ - вероятность отдельного появления каждой основы в корпусе (со сглаженными, то есть защищёнными от деления на ноль, оценками).

Статистическая значимость совместной встречаемости пары проверяется критерием Т:

$$T(a, b) = \frac{f(a, b) - N \times P(a)P(b)}{\sqrt{f(a, b) + 0.5}}. \quad (3)$$

Здесь $f(a, b)$ - наблюдаемая частота пары;

N - общее число рассмотренных контекстных окон корпуса (фрагментов текста из 4 слов).

Показатель Skip суммирует встречи пары со вставленными между её элементами словами с весом $1/(g+1)$, где g - число промежуточных слов; это упрощённый вариант идеи k-skip-n-грамм. Dist - фактическое расстояние между словами кандидата. Order - признак инверсии, принимающий значение 1, если обратный порядок пары встречается в корпусе чаще прямого; этот признак служит для выявления случаев перестановки компонентов фразеологизма.

Модель А - гибрид байесовского подхода и градиентного бустинга. Модель А основана на алгоритме NGBoost. Для каждого кандидата модель предсказывает не один балл, а два параметра нормального распределения - условное среднее $m(x_c)$ (степень фразеологичности) и условную дисперсию $s^2(x_c)$ (дисперсию модели относительно данной оценки). Оба параметра выражаются суммой ансамбля деревьев:

$$m(x_c) = \sum_{m=1}^M f_m(x_c), \quad s^2(x_c) = \exp\left(\sum_{m=1}^M g_m(x_c)\right). \quad (4)$$

Здесь $m(x_c)$ - балл фразеологичности кандидата (условное среднее значение);

$s^2(x_c)$ - дисперсия модели относительно данной ею оценки;

f_m , g_m - поочередно обучаемые регрессионные деревья;

M - число деревьев (в эксперименте 400).

Модель обучается путём минимизации функции потерь отрицательного логарифмического правдоподобия (L_{NLL} - Negative Log-Likelihood):

$$L_{NLL} = \frac{1}{2} \sum_{c=1}^N \left[\ln(2ps^2(x_c)) + \frac{(y_c - m(x_c))^2}{s^2(x_c)} \right]. \quad (5)$$

Здесь N - общее число кандидатов (наблюдений); остальные обозначения введены выше.

На каждой итерации деревья строятся по следующим градиентам (псевдоостаткам):

$$g_c^m = \frac{\nabla L_{NLL}}{\nabla m} = \frac{m(\mathbf{x}_c) - y_c}{s^2(\mathbf{x}_c)}, \quad g_c^s = \frac{\nabla L_{NLL}}{\nabla \ln s^2} = \frac{1}{2} \frac{\partial}{\partial \ln s^2} - \frac{(y_c - m(\mathbf{x}_c))^2}{s^2(\mathbf{x}_c)} \frac{\partial}{\partial \ln s^2} \quad (6)$$

Здесь g_c^m - градиент первого порядка по прогнозируемому значению модели (m);

g_c^s - второй градиент, вычисляемый по логарифмической дисперсии модели ($\ln s^2$).

Первый градиент (g_c^m) направляет прогнозную оценку модели (m) к истинной метке и уменьшает шаг оптимизации в точках, где дисперсия модели относительно данной оценки (s^2) высока. Второй градиент (g_c^s) увеличивает дисперсию модели (s), когда ошибка прогноза выше ожидаемой, и уменьшает её, когда ошибка мала. Именно этот механизм придаёт модели байесовское свойство, то есть способность оценивать границы собственного знания и неопределённости. Модель А обучается только по вектору статистических признаков ($\mathbf{x}_c^{stat}$) и не имеет априорного регуляризующего члена.

Гибридная модель байесовского градиентного бустинга на основе онтологии. Модель В отличается от модели А двумя каналами. Первый канал - расширение пространства признаков. На триплетах онтологии методом TransE обучаются векторные представления (эмбединги). Модель TransE минимизирует граничную функцию потерь (L_{TransE} , Margin Loss), сохраняющую заданное минимальное расстояние (margin) между правильными и искажёнными триплетами:

$$L_{TransE} = \sum_{(h,r,t)} \max(0, g + \mathbf{P} \mathbf{e}_h + \mathbf{e}_r - \mathbf{e}_t \mathbf{P} - \mathbf{P} \mathbf{e}_{h'} + \mathbf{e}_r - \mathbf{e}_t \mathbf{P}). \quad (7)$$

Здесь $\mathbf{e}_h$ и $\mathbf{e}_t$ - эмбединг-векторы соответственно головного узла (h) и хвостового узла (t);

$\mathbf{e}_r$ - эмбединг-вектор отношения (r) между узлами онтологии;

$\mathbf{e}_{h'}$ и $\mathbf{e}_{t'}$ - эмбединг-векторы узлов случайно искажённого (негативно-го/неправильного) триплета;

g - граница разделения (margin);

$\mathbf{P} \times \mathbf{P}$ - норма (длина) вектора.

Если основы кандидата соответствуют шаблону компонентов некоторого фразеологизма в онтологии, эмбединги опорной леммы, семантической категории и структурной категории этого фразеологизма объединяются и образуют онтологический вектор признаков:

$$\mathbf{x}_c^{onto} = [\mathbf{e}_{Anchor} \mathbf{P} \mathbf{e}_{SemCat} \mathbf{P} \mathbf{e}_{StrCat}], \quad \mathbf{x}_c^{hyb} = \frac{1}{2} \mathbf{x}_c^{stat} \mathbf{P} \mathbf{x}_c^{onto} \quad (8)$$

Здесь $\mathbf{P}$ - операция конкатенации векторов;

$\mathbf{x}_c^{onto}$ - онтологический вектор признаков кандидата;

$\mathbf{x}_c^{hyb}$ - итоговый гибридный вектор, объединяющий статистические ($\mathbf{x}_c^{stat}$) и онтологические данные.

Для кандидатов, которым не найден соответствующий фразеологизм, $\mathbf{x}_c^{onto}$ принимается нулевым вектором (размерность эмбединга равна 32, поэтому гибридный вектор имеет в общей сложности 102 измерения).

Второй канал - байесовское априорное распределение. Каждому кандидату до обучения модели назначается нормальный априор:

$$y_c \sim N(m_0(c), s_0^2(c)),$$

$$(m_0, s_0^2) = \begin{cases} (0.90, 0.05), & \text{если онтологические правила выполнены} \\ \text{полностью} \\ (0.50, 0.30), & \text{в противном случае} \end{cases} \quad (9)$$

Полное выполнение онтологических правил включает три условия:

- основы кандидата соответствуют шаблону компонентов некоторого фразеологизма;

- цепочка аффиксов каждого слова полностью сегментируется в пределах допустимых групп аффиксов данной леммы;

- опорная лемма фразеологизма присутствует в предложении.

Если эти правила не выполнены, априорное распределение сохраняется в нейтральном центре ($m_{0,c} = 0.50$), а его дисперсия задаётся высокой. Нейтральный центр отражает лишь отсутствие сведений о сочетании в онтологии, а не отсутствие у него фразеологичности; жёстко низкий центр занижал бы оценки настоящих фразеологизмов, ещё не зафиксированных в онтологии.

На основе заданных выше онтологических правил и априорного распределения гибридная модель (модель В) в процессе обучения добавляет к своей likelihood-части (отрицательному логарифмическому правдоподобию) априорный штраф и минимизирует функцию потерь MAP (байесовский метод оценивания, находящий наиболее вероятную точку апостериорного распределения):

$$L_{MAP} = L_{NLL} + l \sum_{c=1}^N \frac{(m(\mathbf{x}_c^{hyb}) - m_0(c))^2}{s_0^2(c)}. \quad (10)$$

Здесь $m_0(c)$ и $s_0^2(c)$ - априорные среднее и дисперсия кандидата c из формулы (9); остальные обозначения соответствуют формулам (4) и (8).

$l > 0$ - гиперпараметр, определяющий силу влияния онтологии на решение модели; в эксперименте принято $l = 0.3$, поскольку это значение обеспечило наиболее сбалансированный результат. Функция потерь MAP логически состоит из двух частей - её основа (L_{NLL}) обеспечивает соответствие статистическим данным, а добавленный штрафной блок выражает опору модели на априорное знание.

Поскольку штраф квадратичен, в его производной по m появляется коэффициент 2, и градиент обучения деревьев вместе со второй производной (гессианом) записываются следующим образом:

$$\frac{\eta L_{MAP}}{\eta m} = \frac{m - y_c}{s^2} + 2I \frac{m - m_0}{s_0^2}, \quad \frac{\eta^2 L_{MAP}}{\eta m^2} = \frac{1}{s^2} + \frac{2I}{s_0^2}. \quad (11)$$

Реализация выполняется в два этапа. Сначала NGBoost обучает likelihood-часть на гибридных признаках и выдаёт m и s^2 , затем ансамбль LightGBM [13] со специальной целевой функцией на основе градиента и гесса из формулы (11) переобучает m в направлении решения MAP. Апостериорная дисперсия получается по замкнутой формуле:

$$s_{post}^2 = \frac{\frac{1}{\sigma^2} + \frac{2I}{s_0^2}}{\frac{1}{\sigma^2} + \frac{2I}{s_0^2} + 1}. \quad (12)$$

Формула (12) является классическим результатом байесовской теории - апостериорная точность равна сумме точности правдоподобия и априорной точности, поэтому там, где онтология даёт надёжный априор, дисперсия модели уменьшается.

Решающее правило и оценивание. Итоговое решение принимается по нижней границе апостериорного доверительного интервала и определяется следующим образом:

$$LB_c = m(X_c) - z \times s(X_c), \quad (13)$$

$LB_c - q$ – кандидат признаётся фразеологизмом.

Здесь X_c - вектор признаков кандидата c (в модели А - x_c^{stat} , в модели В - x_c^{hyb});

$s(X_c)$ - предсказанное моделью стандартное отклонение (неопределённость);

$z \geq 0$ - гиперпараметр, определяющий строгость доверия;

q - порог класса.

Пара гиперпараметров z и q подбирается в каждом блоке кросс-валидации (фолде) только на обучающей части путём максимизации F1-меры, отражающей баланс точности и полноты модели. Случай $z = 0$ означает, что при принятии решения дисперсия модели не учитывается и используется только оценка $m(X_c)$. Такая гибкость необходима из-за гетероскедастичного характера дисперсии модели.

Для объективного и корректного оценивания эффективности моделей эксперименты проводились методом 5-кратной кросс-валидации с группировкой по предложениям и сохранением пропорций классов. Механизм группировки гарантирует, что все кандидаты, извлечённые из одного предложения, попадают только в один фолд, что предотвращает утечку данных (*data leakage*) между обучающей и тестовой выборками. Для обеспечения объективного сравнения модели А и В тестировались на одинаково сформированных фолдах. Поскольку ансамбли градиентного бустинга содержат стохастические элементы (подвыборку наблюдений и признаков в NGBoost, случайность построения деревьев в LightGBM), весь протокол был повторён с пятью различными начальными значениями генератора случайных чисел

(seed = 7, 42, 123, 555, 2024), управляющими как разбиением на фолды, так и инициализацией моделей; для LightGBM включён детерминированный режим. В таблице 1 приведены значения, усреднённые по пяти запускам.

На уровне кандидатов качество оценивалось по критериям Precision (точность), Recall (полнота) и их гармоническому среднему - F_1 -мере:

$$P = \frac{TP}{TP + FP}, \quad R = \frac{TP}{TP + FN}, \quad F_1 = \frac{2PR}{P + R}. \quad (14)$$

Здесь TP , FP , FN - соответственно число истинно положительных, ложноположительных и пропущенных случаев.

Способность модели отличать настоящие фразеологизмы от обычных словосочетаний оценивалась критерием ROC-AUC. На резко несбалансированных данных стандартная Ассигасу непригодна - модель, слепо отвечающая «не фразеологизм» на все кандидаты, достигла бы 97% верных ответов, не выявив ни одного фразеологизма. Поэтому в качестве основных показателей оценивания на уровне кандидатов выбраны F_1 -мера (вместе с Precision и Recall) и ROC-AUC. При этом ROC-AUC вычисляется по итоговому решающему баллу LB_c из формулы (13); ранжирующая способность по чистому баллу m отдельно рассматривается в разделе «Результаты».

Кроме того, проведено практически важное оценивание на уровне предложений - балл предложения принимался равным максимальному баллу кандидата в нём, и для задачи «присутствует ли фразеологизм в данном предложении?» вычислялись F_1 и AUC.

Результаты

Результаты, усреднённые по пяти запускам 5-кратной кросс-валидации, приведены в табл. 1 (после знака $\pm$ указано стандартное отклонение по запускам).

Таблица 1

Показатели моделей А и В (среднее по пяти запускам 5-кратной кросс-валидации)

Показатель	Модель А (Байес+GB)	Модель В (Байес+GB+Онтология)
Precision	0,242 $\pm$ 0,006	0,268 $\pm$ 0,028
Recall	0,199 $\pm$ 0,016	0,269 $\pm$ 0,012
F1	0,210 $\pm$ 0,010	0,263 $\pm$ 0,018
ROC-AUC	0,727 $\pm$ 0,021	0,744 $\pm$ 0,014
F1 (уровень предложений)	0,509 $\pm$ 0,022	0,585 $\pm$ 0,013
AUC (уровень предложений)	0,615 $\pm$ 0,011	0,662 $\pm$ 0,012

По всем шести усреднённым показателям превосходство на стороне модели В, причём по пяти из них (Precision, Recall, F1 и обоим показателям уровня предложений) оно воспроизводится во всех пяти запусках, а по ROC-AUC - в четырёх запусках из пяти. Наибольший абсолютный рост наблюдался на уровне предложений - F1 поднялся с 0,509 до 0,585, что составляет 7,6 процентного пункта, или 14,9% в относительном выражении. Полнота (Recall) выросла с 0,199 до 0,269 (на 35,2% в относительном выражении), F1 на уровне кандидатов - с 0,210 до 0,263, точность (Precision) - с 0,242 до

0,268. Отдельного внимания заслуживает способность чистого ранжирования - ROC-AUC, вычисленный по баллу m без учёта неопределённости в решающем правиле, у двух моделей практически совпадает (0,788 против 0,793). Таким образом, выигрыш модели В связан прежде всего с более полным охватом фразеологизмов и более качественным итоговым решающим правилом, а не с переупорядочиванием кандидатов.

Заслуживают внимания и результаты отдельного оценивания эффективности априорных онтологических знаний. Число кандидатов, для которых полностью выполнились три онтологических правила и был назначен сильный априор, составило 451, и 39,0% из них подтвердились как настоящие фразеологизмы. Напротив, среди остальных кандидатов с нейтральным априором доля настоящих фразеологизмов составила всего 2,2%. Этот результат показывает, что онтологические правила выделяют кандидатов с настоящими фразеологизмами с плотностью в 18 раз выше.

Анализ неопределённости отражён на рис. 1 (рисунок построен по базовому запуску с $\text{seed} = 42$). В результате подключения онтологии средняя неопределённость модели S по всем кандидатам снизилась на 27,6% (в остальных запусках - в диапазоне 26-28%); в группе настоящих фразеологизмов снижение составило 15,9%, в группе кандидатов с сильным априором - 6,5%, а в группе фразеологизмов со вставленными между компонентами словами неопределённость практически не изменилась (-2,8%, в разных запусках от -3% до +3%). Снижение в первых группах согласуется с механизмом формулы (12), поскольку добавление априорного члена уменьшает апостериорную дисперсию относительно дисперсии правдоподобия. Относительное снижение оказалось наименьшим в группе с сильным априором - эта группа составляет всего 2,1% общего фонда и состоит из сложных для оценивания кандидатов, средняя неопределённость которых в обеих моделях остаётся выше общего показателя (рис. 1). Выявление разрывных употреблений со вставленными словами остаётся открытым ограничением подхода. В целом модель В принимает свои решения в среднем с меньшей неопределённостью.

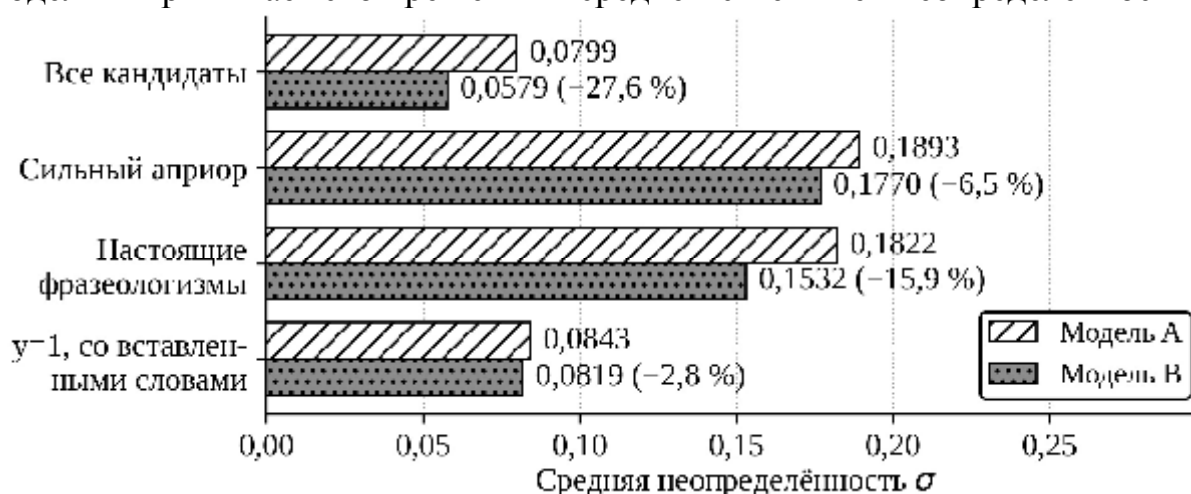


Рис. 1 - Средняя неопределённость σ по группам кандидатов (базовый запуск, $\text{seed} = 42$)

Обсуждение

Важнейший практический результат исследования проявился в оценивании на уровне предложений. Главной прикладной целью моделирования было правильное распознавание предложений, содержащих фразеологизмы, и именно в этом измерении введение онтологии показало свою высокую эффективность (показатель F_1 на уровне предложений поднялся с 0,509 до 0,585). Логическая причина в том, что априорное онтологическое знание повышает оценку самого надёжного сочетания-кандидата в составе предложения. Поскольку общий балл предложения определяется максимальным баллом кандидата в нём, одно точное онтологическое совпадение обеспечивает перевод всего предложения в правильный класс. Данный рост имеет непосредственную практическую ценность для лексикографии и машинного перевода.

Второе направление относится к росту Recall. Повышение полноты с 0,199 до 0,269 показывает, что онтология помогает модели находить сложные случаи, ранее пропускавшиеся, то есть фразеологизмы с изменённой за счёт аффиксов формой или со вставленными посторонними словами. Механизм ясен - правило `allowedAffixGroup` связывает морфологически изменённую форму слова с леммой, а соответствие шаблону компонентов проверяется с расширенным допуском по расстоянию, поэтому статистически редкие, но онтологически естественные формы также получают сильный априор. Это и есть решение проблем морфологической изменчивости и разрывности.

Показатель Precision (точность) у модели В в среднем также выше - 0,268 против 0,242, и это превосходство воспроизвелось во всех пяти запусках. Вместе с тем разброс точности у модели В заметно больше ($\pm 0,028$ против $\pm 0,006$ у модели А), то есть в отдельных запусках значения двух моделей практически совпадают. Повышение F_1 -меры с 0,210 до 0,263 показывает, что рост полноты достигнут без потери точности.

Исходя из требований прикладного применения, баланс между точностью и полнотой можно настраивать с помощью гиперпараметра l и порога класса q ; формулы (10) и (13) обеспечивают эти гибкие механизмы управления. Кроме того, требование научной объективности не позволяет опираться на единственный запуск стохастического ансамбля - при повторных запусках с разными инициализациями отдельные показатели модели В колебались заметно (например, ROC-AUC по решающему баллу от 0,729 до 0,795). Именно поэтому итоговые значения усреднены по пяти запускам, а стандартные отклонения приведены в табл. 1. По ROC-AUC модель В превзошла модель А в четырёх запусках из пяти; превосходство не гарантируется в каждом отдельном запуске, что объясняется объёмом набора данных из 860 предложений и стохастической природой ансамблей.

Следует также объективно отметить невысокий абсолютный уровень значений F_1 на уровне кандидатов (0,210 и 0,263). Этому способствуют три фактора. Первый - резкая несбалансированность классов, то есть лишь 3,0% из 21 084 кандидатов относятся к положительному классу; второй - ограни-

ченный объём золотого стандарта (860 предложений); третий - сама задача, то есть чисто автоматическое различение свободного словосочетания от фразеологизма является одной из сложнейших задач лингвистики. В этих условиях относительное сравнение двух моделей на одинаковых фолдах и показатели уровня предложений дают более надёжную основу для практических выводов.

Снижение неопределённости - чисто байесовское преимущество, доступное только модели В за счёт внешнего знания. На практике выход модели В может использоваться в режиме с управляемым уровнем доверия по правилу формулы (13) - например, составитель словаря может просматривать только кандидатов с высокой нижней границей.

С методологической точки зрения сформированы ещё два важных практических вывода - правильный выбор априорного центра и превращение параметра z решающего правила в настраиваемый гиперпараметр. В первоначальном варианте для кандидатов вне онтологических правил рассматривался низкий априорный центр ($m_{0,c} = 0.10$), однако сравнительные испытания показали устойчивое превосходство нейтрального центра ($m_{0,c} = 0.50$) по ROC-AUC (в среднем на +0,05). Этот результат - эмпирическое подтверждение принципа *«отсутствие в онтологии не означает отсутствие фразеологичности»*, важного при проектировании априорных систем, основанных на знаниях.

Второе наблюдение - превращение параметра z решающего правила в гиперпараметр. Жёсткое значение z при гетероскедастичной дисперсии нарушает ранжирование кандидатов - настоящие фразеологизмы с высоким баллом m нередко имеют и высокую неопределённость, и большое z необоснованно занижает их оценку; подбор z на обучающей части каждого фолда устраняет этот эффект.

Заключение

В статье предложена модель, присоединяющая онтологические знания к гибриду байесовского подхода и градиентного бустинга для извлечения узбекских фразеологизмов из текста по двум каналам - через векторные представления (эмбединги) узлов онтологии и через начальное вероятностное (MAP-априорное) распределение.

Согласно результатам, усреднённым по пяти независимым запускам, обогащённая онтологией модель показала превосходство по всем шести показателям. Важнейший с практической точки зрения показатель F_1 на уровне предложений вырос с 0,509 до 0,585, полнота (Recall) - с 0,199 до 0,269, F_1 на уровне кандидатов - с 0,210 до 0,263; неопределённость модели при принятии решений (сигма) снизилась в среднем на 27%. Доля настоящих фразеологизмов среди кандидатов, получивших сильное априорное знание по онтологическим правилам, оказалась в 18 раз выше, чем среди остальных. Эти результаты доказывают, что предложенная методология обладает высокой практической и измеримой эффективностью для языков с ограниченными ресурсами.

Список использованных источников

1. Sag I. A., Baldwin T., Bond F., Copestake A., Flickinger D. Multiword Expressions: A Pain in the Neck for NLP// Lecture Notes in Computer Science, 2002, vol. 2276. DOI: 10.1007/3-540-45715-1_1.
2. Constant M. et al. Multiword Expression Processing: A Survey// Computational Linguistics. 2017. Vol. 43, No. 4. P. 837-892.
3. Salaev U. UzMorphAnalyser: A Morphological Analysis Model for the Uzbek Language Using Inflectional Endings// AIP Conference Proceedings. 2024. Vol. 3244, No. 1. P. 030058. DOI: 10.1063/5.0241461.
4. Abdurakhmonova N., Ismailov A., Sayfulleyeva R. MorphUz: Morphological Analyzer for the Uzbek Language// 2022 7th Int.Conf. on Computer Science and Engineering (UBMK). - IEEE, 2022. P. 61-66. DOI: 10.1109/UBMK55850.2022.9919579.
5. Nazirova E., Usmonova K. Algorithmic Strategies of Complex Word Analyzing from Uzbek to English for Machine Translation// AIP Conference Proceedings. 2024. Vol. 3244, No. 1. P. 030054. DOI: 10.1063/5.0242104.
6. Nazirova E.Sh. et al. UzbNERCorpus: Developing An Annotation Scheme For NER Corpus In The Uzbek Language// Proc. of the 2026 IEEE Ural-Siberian Conf. on Biomedical Engineering, Radioelectronics and Information Technology (USBREIT). - IEEE, 2026. P. 1-5. DOI: 10.1109/USBREIT70063.2026.11580709.
7. Church K.W., Hanks P. Word Association Norms, Mutual Information, and Lexicography// Computational Linguistics. 1990. Vol. 16, No. 1. P. 22-29.
8. Friedman J.H. Greedy Function Approximation: A Gradient Boosting Machine// The Annals of Statistics. 2001. Vol. 29, No. 5. P. 1189-1232.
9. Duan T. et al. NGBoost: Natural Gradient Boosting for Probabilistic Prediction //Proc. of the 37th Int. Conf. on Machine Learning (ICML 2020). PMLR. 2020. Vol. 119. P. 2690-2700.
10. Gruber T.R. A Translation Approach to Portable Ontology Specifications// Knowledge Acquisition. 1993. Vol. 5, No. 2. P. 199-220.
11. Bordes A. et al. Translating Embeddings for Modeling Multi-relational Data// Advances in Neural Information Processing Systems 26 (NeurIPS 2013). - Red Hook, NY: Curran Associates, 2013. P. 2787-2795.
12. Nazirova E., Abidova Sh., Xolmuminov Y., Boymurodov F., Madetbayeva N. Mathematical Model and Algorithm for Grammatical Editing of Given Words in Uzbek Language // International Seminar on Mathematics for Sustainable Industry 2024. - Kuala Lumpur, 2025. P. 175-182.
13. Ke G. et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree // Advances in Neural Information Processing Systems 30 (NeurIPS 2017). - Red Hook, NY: Curran Associates, 2017. P. 3146-3154.

Назирова Э., Бахромова Л.

**ЭКОНОМИКО-МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ОПРЕДЕЛЕНИЯ
ЧИСЛЕННОСТИ ОФИЦИАЛЬНОГО И НЕФОРМАЛЬНОГО
ПЕРСОНАЛА**

**Ташкентский университет информационных технологий имени
Мухаммада аль-Хорезми, Ташкент, Узбекистан**

1. Введение

Высокая доля неформальной занятости - это устойчивая черта экономик развивающихся стран, охватывающая большую часть рабочей силы. Такая деятельность не регистрируется официально, что создает серьезные трудности для государственного управления, планирования и социальной политики.

Понятие неформальной экономики впервые ввел Харт [1]. Международная организация труда [2] уточнила, что нужно различать неформальный сектор и неформальную занятость. Главное здесь, отсутствие правовой и социальной защиты, а не то, в какой отрасли человек работает.

Моделирование динамики рынков труда развивалось по нескольким направлениям. Классические подходы использовали балансовые и матричные модели для описания перетока рабочей силы между секторами [3, 4]. Более поздние исследования опирались на системы дифференциальных уравнений, позволяющие рассматривать процесс в непрерывном времени [5-7]. Однако в большинстве работ неформальная занятость рассматривается как остаточная величина [8-13]. Внутренняя структура неформальной занятости, а именно различие между официально зарегистрированными самозанятыми и теми, кто работает без всякой регистрации, обычно не учитывается. Именно здесь существует пробел в исследованиях.

В литературе не хватает динамической модели, которая учитывала бы структуру занятости с выделением отдельных секторов, описывала бы переходы между ними в непрерывном времени и при этом обладала бы строгими математическими свойствами. Такие свойства позволили бы анализировать устойчивость и ставить задачи оптимального управления. Настоящая работа призвана восполнить этот пробел. Цель исследования, построить динамическую модель структуры занятости, доказать ее математическую корректность и определить условия, при которых доля неформального сектора стабильно уменьшается.

2. Математическая модель блока демографической динамики

Вся рабочая сила разделена на фонды, которые непрерывно меняются со временем. Общая численность экономически активного населения обозначена как $N_{\text{tot}}(t)$. Выделены следующие секторы: теневой сектор (N_{shadow}), зарегистрированные самозанятые (N_{self}), корпоративный сектор (N_{corp}), бюджетный сектор (N_{budget}) и индивидуальные предприниматели (N_{pred}). Поскольку численность людей в каждом секторе зависит от времени, скорость ее изменения описывается частной производной $\partial N / \partial t$. Система имеет вид:

$$\begin{aligned} \frac{\partial N_{\text{shadow}}}{\partial t} &= g_{\text{pop}}(t) N_{\text{shadow}} - k(N_{\text{self}}) N_{\text{shadow}} - m_1 N_{\text{shadow}}, \\ \frac{\partial N_{\text{self}}}{\partial t} &= g_{\text{pop}}(t) N_{\text{self}} + k(N_{\text{self}}) N_{\text{shadow}} - m_2 N_{\text{self}}, \\ \frac{\partial N_{\text{corp}}}{\partial t} &= g_{\text{pop}}(t) N_{\text{corp}} + m_1 N_{\text{shadow}}, \\ \frac{\partial N_{\text{budget}}}{\partial t} &= g_{\text{pop}}(t) N_{\text{budget}}, \\ \frac{\partial N_{\text{pred}}}{\partial t} &= g_{\text{pop}}(t) N_{\text{pred}} + m_2 N_{\text{self}}. \end{aligned} \quad (1)$$

t - непрерывное время (в годах).

$N_{\text{tot}}(t)$ - общая численность экономически активного населения;

$N_{\text{shadow}}(t)$ - численность занятых в теневом (неформальном) секторе;

$N_{\text{self}}(t)$ - численность зарегистрированных самозанятых;

$N_{\text{corp}}(t)$ - численность занятых в корпоративном секторе;

$N_{\text{budget}}(t)$ - численность занятых в бюджетном секторе;

$N_{\text{pred}}(t)$ - численность индивидуальных предпринимателей;

$g_{\text{pop}}(t)$ - естественный непрерывный темп прироста трудоспособного населения;

$k(t)$ - интенсивность выхода из теневого сектора в категорию самозанятых (детеневилизация);

$N(t) = (N_{\text{shadow}}(t), N_{\text{self}}(t), N_{\text{corp}}(t), N_{\text{budget}}(t), N_{\text{pred}}(t))^T$ - коэффициент перетока из теневого сектора в корпоративный сектор;

$m_2(t)$ - коэффициент перехода самозанятых в индивидуальные предприниматели.

Интенсивность формализации зависит не от времени, а от текущего состояния системы. Когда число зарегистрированных самозанятых приближается к предельной вместимости K , темп новой формализации замедляется. Это отражает эффект насыщения,

$$k(N_{\text{self}}) = k_0 \frac{\partial}{\partial N_{\text{self}}} \left(1 - \frac{N_{\text{self}}}{K} \right) \ddot{} k_0 > 0, K > 0. \quad (2)$$

Благодаря этому в первых двух уравнениях появляется произведение $N_{\text{self}} \times N_{\text{shadow}}$, то есть нелинейная связь, что придаёт модели содержательный нелинейный характер.

Условие сохранения баланса. Если сложить все пять уравнений, то члены, описывающие переходы между секторами, попарно сократятся. Останется следующее соотношение.

$$\frac{\mathbb{I} N_{\text{tot}}}{\mathbb{I} t} = \mathring{\mathbb{A}}_m \frac{\mathbb{I} N_m}{\mathbb{I} t} = g_{\text{pop}}(t) N_{\text{tot}}(t), N_{\text{tot}}(t) = N_{\text{tot}}(0) e^{g_{\text{pop}} t} \quad (3)$$

Таким образом, общая численность занятых следует демографическому тренду, независимо от того, насколько активно происходят переходы между секторами. Это равенство служит аналитической проверкой корректности модели. Представленная система описывает, как перераспределяется трудоспособное население между различными секторами экономики с течением времени t . Несмотря на использование в записи символа частной производной $\mathbb{I} N / \mathbb{I} t$, все неизвестные функции зависят только от одной переменной - времени t .

С математической точки зрения это система обыкновенных дифференциальных уравнений (ОДУ) первого порядка. Нелинейность в системе возникает из-за функции интенсивности формализации $k(N_{\text{self}})$. Ниже представ-

лена постановка условий и пошаговый алгоритм численного решения этой системы методом конечных разностей.

В данной математической модели пространственные граничные условия не задаются, поскольку в системе отсутствуют пространственные переменные, такие как координаты x, y, z . Процесс рассматривается как сосредоточенный, представляющий собой глобальный макроэкономический баланс. Таким образом, задача формулируется как задача Коши, требующая задания начальных условий. Для однозначного решения необходимо определить численность населения в каждом из пяти секторов в начальный момент времени $t = 0$.

$$\begin{aligned} N_{\text{shadow}}(0) &= N_{\text{shadow}}^0, N_{\text{self}}(0) = N_{\text{self}}^0, N_{\text{corp}}(0) = N_{\text{corp}}^0, \\ N_{\text{budget}}(0) &= N_{\text{budget}}^0, N_{\text{pred}}(0) = N_{\text{pred}}^0. \end{aligned} \quad (4)$$

Сумма этих начальных значений определяет базовую общую численность экономически активного населения.

$$N_{\text{tot}}(0) = N_{\text{shadow}}^0 + N_{\text{self}}^0 + N_{\text{corp}}^0 + N_{\text{budget}}^0 + N_{\text{pred}}^0. \quad (5)$$

2.1. Корректность задачи Коши и условие сокращения теневого сектора

Обозначим $N(t) = (N_{\text{shadow}}(t), N_{\text{self}}(t), N_{\text{corp}}(t), N_{\text{budget}}(t), N_{\text{pred}}(t))^T$ и запишем систему (1) в векторной форме.

$$\frac{dN}{dt} = F(t, N), N(0) = N^0, \quad (6)$$

где правая часть $F(t, N)$ определяется соотношениями (1)-(2), а начальный вектор N^0 - условиями (4).

Утверждение 1 (существование и единственность решения). Пусть функция $g_{\text{pop}}(t)$ непрерывна на отрезке $[0, T]$, параметры k_0, K, m_1, m_2 постоянны, причём $k_0 > 0$, $K > 0$, $m_1 \geq 0$, $m_2 \geq 0$, а начальные данные (4) неотрицательны. Тогда задача Коши (6) имеет единственное решение, определённое на всём отрезке $[0, T]$, и это решение неотрицательно.

Доказательство. Компоненты вектор-функции $F(t, N)$ являются многочленами не выше второй степени относительно переменных N с непрерывно зависящими от t коэффициентами. Единственный нелинейный член $k(N_{\text{self}})N_{\text{shadow}} = k_0(1 - N_{\text{self}}/K)N_{\text{shadow}}$ билинеен. Следовательно, функция F непрерывна по совокупности переменных и непрерывно дифференцируема по N , а значит, локально липшицева по N . По теореме Пикара-Линделёфа [14] решение задачи (6) существует и единственно в некоторой окрестности точки $t = 0$.

Покажем инвариантность неотрицательного октанта. На грани $N_{\text{shadow}} = 0$ имеем $dN_{\text{shadow}}/dt = 0$; на грани $N_{\text{self}} = 0$ получаем $dN_{\text{self}}/dt = k_0 N_{\text{shadow}} \geq 0$; на гранях $N_{\text{corp}} = 0$ и $N_{\text{pred}} = 0$ соответственно $dN_{\text{corp}}/dt = m_1 N_{\text{shadow}} \geq 0$ и $dN_{\text{pred}}/dt = m_2 N_{\text{self}} \geq 0$; на грани $N_{\text{budget}} = 0$ имеем

$dN_{\text{budget}} / dt = 0$. Таким образом, при неотрицательных начальных данных траектория не покидает область $N \geq 0$.

В силу закона сохранения (3) сумма компонент удовлетворяет равенству $N_{\text{tot}}(t) = N_{\text{tot}}(0)e^{\int_0^t g_{\text{pop}}(s) ds}$, откуда для всех $t \in [0, T]$ справедлива оценка

$$0 \leq N_i(t) \leq N_{\text{tot}}(t) \leq M, \quad M = N_{\text{tot}}(0)e^{\int_0^T \max_{[0, T]} |g_{\text{pop}}| ds}. \quad (7)$$

Решение остаётся ограниченным на любом конечном интервале, поэтому разрушение решения за конечное время исключено и локальное решение продолжается на весь отрезок $[0, T]$. На компактном множестве

$W = \{N \geq 0 : \sum_i N_i \leq M\}$ матрица Якоби $\|F / \|N$ ограничена, что даёт равномерную постоянную Липшица $L = L(M, k_0, K, m_1, m_2, g_{\text{pop}})$ и обеспечивает единственность решения на всём отрезке. Утверждение доказано.

Отметим, что оценка (7) одновременно служит теоретическим обоснованием контрольного соотношения, применяемого в разностной схеме раздела 3.

Утверждение 2 (условие устойчивого сокращения теневого сектора). Первое уравнение системы (1) линейно относительно N_{shadow} , что позволяет выписать его решение в явном виде.

$$N_{\text{shadow}}(t) = N_{\text{shadow}}(0)e^{\int_0^t (g_{\text{pop}}(s) - k(N_{\text{self}}(s)) - m_1) ds}. \quad (8)$$

Для доли теневого сектора $a_{\text{shadow}}(t) = N_{\text{shadow}}(t) / N_{\text{tot}}(t)$ с учётом закона сохранения (3) получаем следующее уравнение.

$$\frac{da_{\text{shadow}}}{dt} = - \frac{m_1}{K} (N_{\text{self}}(t)) + m_2 a_{\text{shadow}}(t). \quad (9)$$

Отсюда доля теневого сектора монотонно убывает тогда и только тогда, когда $k(N_{\text{self}}) + m_1 > 0$, то есть при выполнении условия

$$N_{\text{self}}(t) < K \frac{m_1}{m_2 + m_1}. \quad (10)$$

Условие (10) не зависит от темпа демографического прироста g_{pop} : даже при быстром росте населения доля теневого сектора сокращается, если численность зарегистрированных самозанятых не превышает порогового значения $K(1 + m_1/k_0)$. Для параметров расчёта раздела 4 ($K = 200$, $m_1 = 0,1$, $k_0 = 0,5$) это пороговое значение равно 240 тыс. человек, тогда как максимальное расчётное значение N_{self} составляет 195 тыс. человек. Следовательно, на всём горизонте моделирования доля теневого сектора убывает, что подтверждается данными табл. 1.

3. Численная схема и алгоритм решения

Для решения системы на компьютере непрерывная модель переводится в дискретную. Будем использовать явную разностную схему первого порядка точности [14-16]. Вводится равномерная временная сетка с шагом интегри-

рования $Dt > 0$. Приближенные значения функций ищутся в узлах сетки $t_n = n \times Dt$, где $n = 0, 1, 2, \dots$ - номер временного шага. Приближенное значение искомой функции на n -м шаге обозначается верхним индексом.

$$t_n = nDt, n = 0, 1, 2, \dots, N_{shadow}^n \gg N_{shadow}(t_n). \quad (11)$$

Непрерывные скорости изменения численности dN/dt заменяются разностными производными.

$$\frac{dN}{dt} \gg \frac{N^{n+1} - N^n}{Dt}. \quad (12)$$

Разностные аппроксимации подставляются в исходную систему уравнений. Правые части уравнений вычисляются на текущем слое n . Сначала на каждом шаге n рассчитывается текущая интенсивность выхода из теневого сектора, зависящая от состояния системы.

$$k^n = k_0 \left(1 - \frac{N_{self}^n}{K} \right), k_0 > 0, K > 0. \quad (13)$$

Затем вычисляются значения для каждого сектора на следующем шаге $n+1$.

$$\begin{aligned} N_{shadow}^{n+1} &= N_{shadow}^n + Dt \left(g_{pop}(t_n) N_{shadow}^n - k^n N_{shadow}^n - m_1 N_{shadow}^n \right), \\ N_{self}^{n+1} &= N_{self}^n + Dt \left(g_{pop}(t_n) N_{self}^n + k^n N_{shadow}^n - m_2 N_{self}^n \right), \\ N_{corp}^{n+1} &= N_{corp}^n + Dt \left(g_{pop}(t_n) N_{corp}^n + m_1 N_{shadow}^n \right), \\ N_{budget}^{n+1} &= N_{budget}^n + Dt \left(g_{pop}(t_n) N_{budget}^n \right), \\ N_{pred}^{n+1} &= N_{pred}^n + Dt \left(g_{pop}(t_n) N_{pred}^n + m_2 N_{self}^n \right). \end{aligned} \quad (14)$$

Пять уравнений описывают динамику теневого сектора, зарегистрированных самозанятых, корпоративного сектора, бюджетного сектора и индивидуальных предпринимателей соответственно.

Алгоритмическая реализация и контроль устойчивости. Вычисления выполняются циклически: значения, полученные на слое $n+1$, становятся исходными данными для расчета слоя $n+2$ и так далее до достижения конечного времени моделирования. Учитывая, что в первых двух уравнениях присутствует нелинейная зависимость в виде произведения $N_{self} \times N_{shadow}$ (внутри функции k), явная разностная схема может быть условно устойчивой. Это требует выбора достаточно малого шага по времени Dt (согласно критерию Куранта-Фридрихса-Леви для обыкновенных дифференциальных уравнений), чтобы избежать расходимости алгоритма. На каждом расчетном шаге рекомендуется применять аналитический критерий проверки корректности модели.

$$N_{tot}^{n+1} = \sum_m N_m^{n+1}. \quad (15)$$

Суммарное значение должно удовлетворять дискретному аналогу уравнения баланса.

$$N_{\text{tot}}^{n+1} \gg N_{\text{tot}}^n (1 + Dt g_{\text{pop}}(t_n)). \quad (16)$$

Если вычислительная ошибка нарастает и баланс нарушается, следует либо уменьшить шаг Dt , либо перейти к использованию неявных разностных схем, которые требуют применения итерационных методов (например, метода Ньютона) на каждом временном шаге.

Блок-схема алгоритма решения. На рис. 1 представлена структурная схема алгоритма численного решения модели методом конечных разностей. Схема показывает последовательность вычислительных этапов и организацию итерационного цикла по времени, обеспечивающего пошаговое построение траектории всех пяти секторов.

Обоснование и комментарии к алгоритму. Блок 4 вынесен перед основным расчетом, что критически важно. Интенсивность формализации зависит от текущего состояния системы $k(N_{\text{self}})$, поэтому она должна пересчитываться на каждом шаге интегрирования до расчета потоков между секторами. Блок 6 (верификация) служит встроенным аналитическим критерием проверки корректности модели. Если численность в секторах рассчитана с ошибкой (например, из-за слишком большого шага Dt), сумма перестанет соответствовать экспоненциальному закону, и алгоритм сможет выдать предупреждение.

$$N_{\text{tot}}(t) = N_{\text{tot}}(0) e^{g_{\text{pop}} t}. \quad (17)$$

4. Результаты математического моделирования

Для наглядности смоделируем ситуацию с конкретными параметрами. Предположим, демографический прирост отсутствует ($g_{\text{pop}} = 0$), чтобы наглядно видеть чистый переток людей.

Параметры: $k_0 = 0,5$; $m_1 = 0,1$; $m_2 = 0,1$; ёмкость самозанятых $K = 200$.

Начальные условия:

теневого сектор = 400,

самозанятые = 50,

корпоративный = 300,

бюджетный = 200,

ИП = 50.

Общая численность $N_{\text{tot}} = 1000$. Шаг $Dt = 1$ год. Результаты расчёта приведены в табл. 1.

В первые годы происходит резкий скачок легализации, самозанятые растут с 50 до 195. Затем коэффициент k падает почти до нуля, поскольку приближается к предельной емкости K , что отражает эффект насыщения. Суммарное значение на каждом шаге строго остается равным 1000, что подтверждает корректность разностной схемы.

Динамика пяти процессов. На рис. 2 представлена визуализация решения системы на десятилетнем горизонте. Пять кривых показывают численность секторов во времени, а штриховая линия отражает суммарную численность, служащую проверкой баланса.

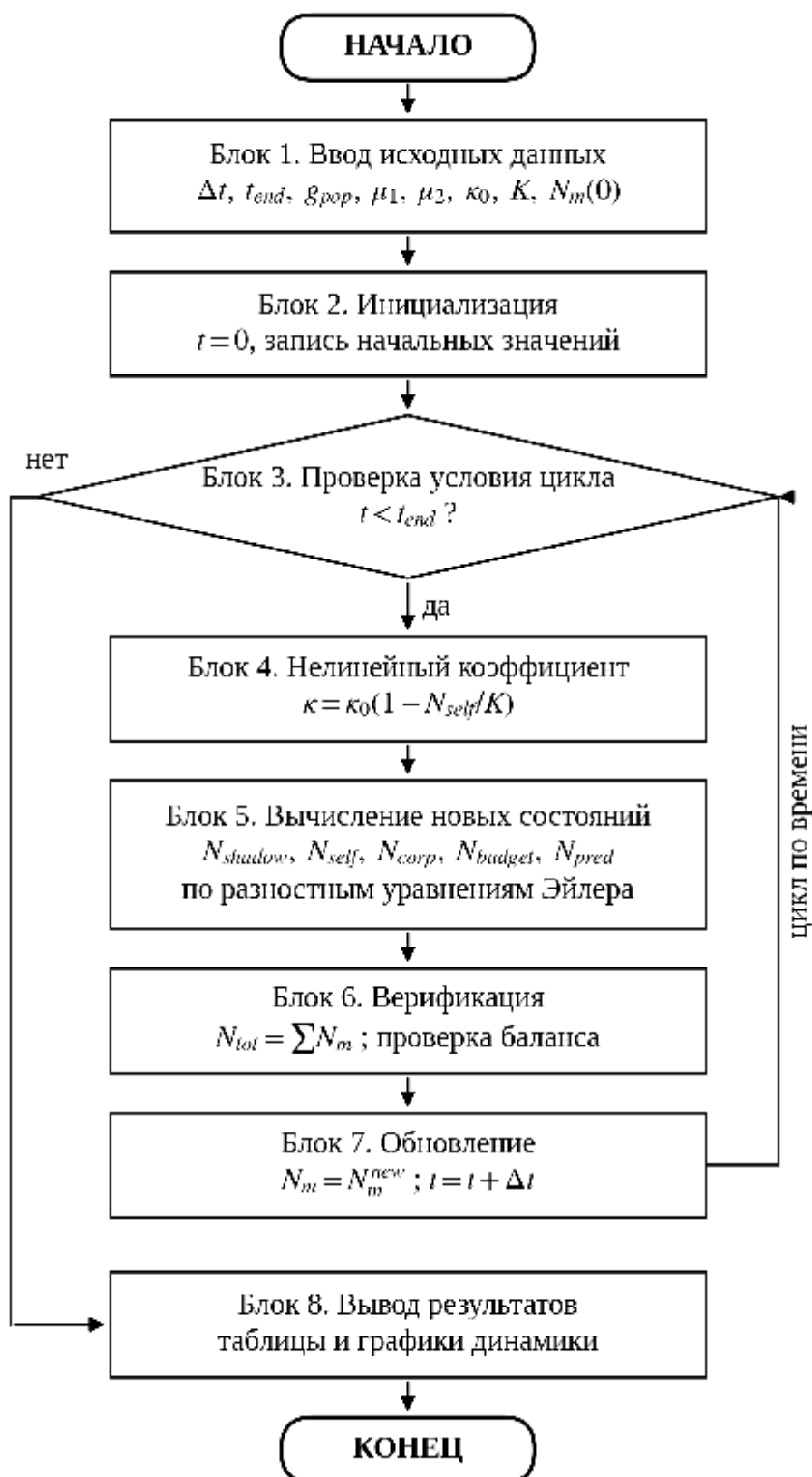


Рис. 1. Блок-схема алгоритма численного решения модели методом конечных разностей

Объяснение динамического состояния. Кривая с круглыми маркерами (теневой сектор). Демонстрирует экспоненциальный спад. Наиболее резкое падение наблюдается в первый год. Численность активно убывает из-за двух

потоков: перехода в корпоративный сектор (коэффициент m_1) и интенсивного выхода в самозанятость (k).

Таблица 1

Численность по секторам, тыс. человек ($Dt = 1$ год, $g_{\text{пор}} = 0$)

Год	Теневой	Самозан.	Корпор.	Бюджет	ИП	Всего
0	400,0	50,0	300,0	200,0	50,0	1000,0
1	210,0	195,0	340,0	200,0	55,0	1000,0
2	186,4	178,1	361,0	200,0	74,5	1000,0
3	157,5	170,5	379,6	200,0	92,3	1000,0
4	130,2	165,1	395,4	200,0	109,4	1000,0
5	105,8	159,9	408,4	200,0	125,9	1000,0

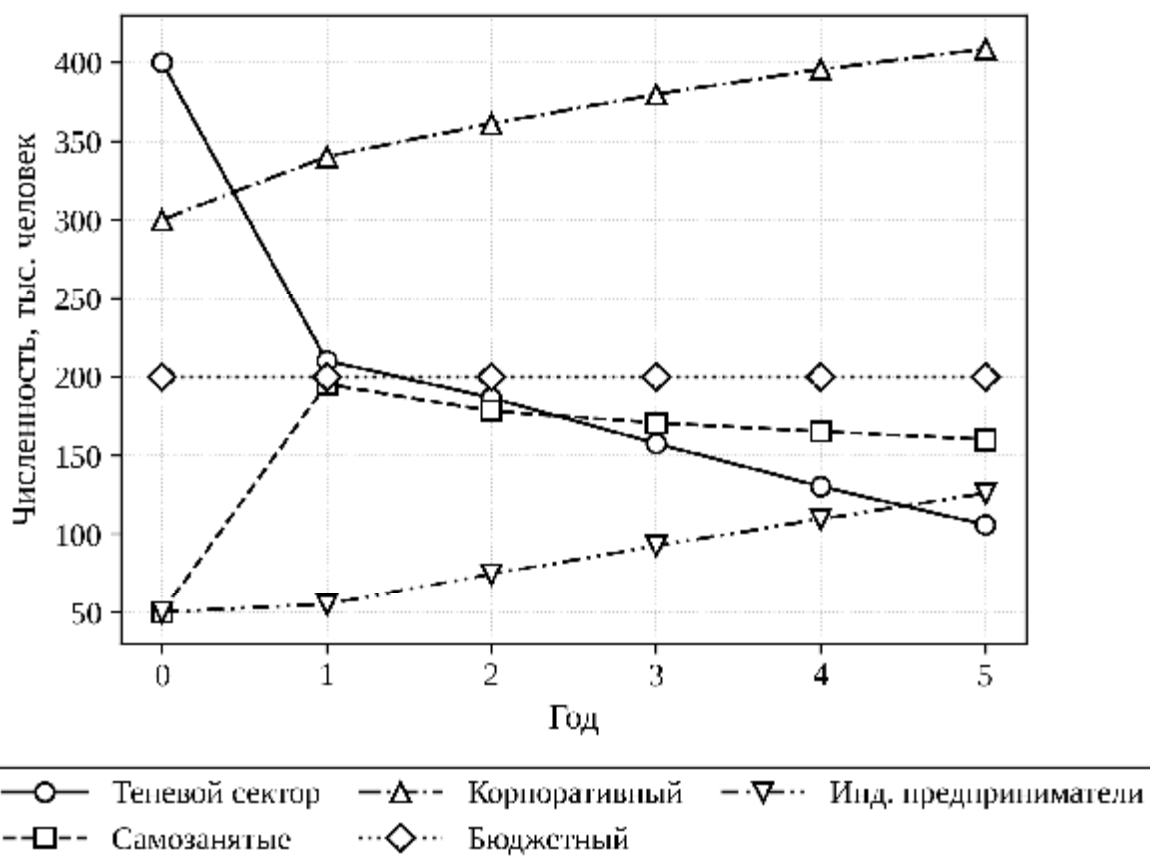


Рис. 2. Динамика численности пяти секторов и линия баланса

Кривая с квадратными маркерами (самозанятые). Имеет ярко выраженный нелинейный характер. В начале виден резкий взлёт (всплеск детеневизации). Однако по мере приближения к отметке 200 (предельная ёмкость K) темп новой формализации снижается, что отражает эффект насыщения. Кривая упирается в потолок и даже слегка проседает, так как люди из самозанятых продолжают переходить в категорию ИП (отток m_2).

Кривые с треугольными маркерами (корпоративный сектор и ИП). Обе показывают стабильный, почти монотонный рост. Они выступают реципиентами в системе, корпоративный сектор пополняется из теневого (m_1), а ИП пополняется из пула самозанятых (m_2).

Пунктирная линия с ромбовидными маркерами (бюджетный сектор). Абсолютно прямая и параллельна оси времени (константа 200). В уравнении для бюджетного сектора отсутствуют потоки обмена с другими фондами, а поскольку общий прирост населения $g_{\text{pop}} = 0$, этот сектор остаётся неизменным. Штриховая линия сверху (N_{tot}). Проверочная линия баланса. На протяжении всей симуляции она строго держится на отметке 1000. Члены переходов взаимно сократились, что доказывает устойчивость разностной схемы и отсутствие утечки людей в процессе численного решения.

Модель описывает перераспределение рабочей силы между секторами экономики. Сжатие теневого сектора: предполагается, что стимулы к легализации приведут к максимальному оттоку из теневой экономики в самом начале. Со временем легализация станет сложнее, а скорость уменьшения теневого сектора будет зависеть от параметров m и $k(N_{\text{self}})$.

Нелинейное насыщение самозанятых: этот сектор наиболее сложен для прогнозирования. Модель учитывает, что резерв для пополнения не бесконечен, основываясь на произведении числа самозанятых (N_{self}) и числа занятых в тени (N_{shadow}). Прогноз показывает, когда программа регистрации самозанятых исчерпает свой потенциал и начнется переход в статус индивидуальных предпринимателей (ИП).

Стабильное поглощение корпоративным сектором: это основной получатель рабочей силы. Прогнозирование его динамики важно для оценки будущих налоговых поступлений и потребности в рабочих местах. Рост в этом секторе обеспечивается перетоком из теневой экономики с коэффициентом m .

Индивидуальные предприниматели как вторичная эволюция: модель прогнозирует, что ИП будут пополняться за счет самозанятых, которые переросли свой текущий статус (коэффициент m_2). Рост числа ИП представляет собой вторую волну, следующую за начальным ростом регистраций самозанятых.

Бюджетный сектор и демографическая инерция: этот сектор либо остаётся стабильным (при $g_{\text{pop}} = 0$), либо следует демографическим трендам. Он изолирован от перетоков рабочей силы, поэтому его прогнозирование зависит только от демографии.

Общая динамика: модель демонстрирует структурные сдвиги. При суммировании всех уравнений внутренние переходы сокращаются, и общая численность экономически активного населения меняется только за счет естественного прироста. Это означает, что рост легальных секторов происходит за счет сокращения теневого сектора. График подтверждает нулевой баланс: площадь под кривой теневого сектора перераспределяется между растущими легальными секторами.

5. Математические свойства модели

Изоморфизм задачам массопереноса: система обыкновенных дифференциальных уравнений моделирует баланс, схожий с моделями многофазных

потоков или фильтрации в сообщающихся резервуарах. Коэффициенты переходов действуют как градиенты, запускающие потоки, а секторы экономики - как узлы вычислительной сетки.

Глобальный инвариант и закон сохранения: наиболее значимое свойство - наличие аналитического условия баланса. Сумма всех уравнений показывает, что внутренние переходы взаимно уничтожаются, что выражает закон сохранения: общая численность населения (N_{tot}) зависит только от внешнего фактора (демографического тренда g_{pop}). Это свойство служит надежным инвариантом для проверки численных методов.

Нелинейность и асимптотическая устойчивость: функция интенсивности детеневизации $k(N_{\text{self}})$, зависящая от состояния самозанятых, делает систему нелинейной. Множитель $(1 - N_{\text{self}} / K)$ делает модель похожей на логистическое уравнение Ферхюльста [17], описывающее ограниченный рост. Математически это исключает сингулярности и обеспечивает стремление системы к квазистационарному состоянию (аттрактору), где потоки стабилизируются.

Вычислительные ограничения: наличие произведения переменных ($N_{\text{self}} \times N_{\text{shadow}}$) в уравнениях накладывает ограничения на численные методы. Явные разностные схемы требуют строгого контроля шага интегрирования для предотвращения неустойчивости. При усложнении модели могут потребоваться неявные схемы и матричные алгоритмы.

6. Заключение

Численное решение дифференциальной модели выявило сложную нелинейную динамику изменений на рынке труда. Анализ пятилетнего периода моделирования позволяет сделать следующие выводы, внедрение специального режима для самозанятых дает первоначальный быстрый эффект, который затем замедляется. Это связано с нелинейной функцией, которая ограничивает приток новых работников в этот сектор по мере приближения к его предельной вместимости.

Режим самозанятости оказался временным этапом. После достижения пика численность самозанятых начала снижаться из-за постоянного перехода в статус ИП. ИП, напротив, демонстрируют отложенный, но стабильный рост, формируя вторую волну формализации.

Корпоративный сектор стал основным бенефициаром. При заданных коэффициентах перехода он абсорбировал наибольшую долю работников, вышедших из тени, и стал крупнейшим сегментом занятости.

Бюджетный сектор оказался нечувствительным к процессам детенивизации. Его численность осталась неизменной, что подтверждает корректность его дифференциального уравнения.

На протяжении всего моделирования общая численность населения оставалась на уровне 1000 человек. Это доказывает, что рост формальной экономики происходит за счет перераспределения внутренних ресурсов, а не внешнего притока.

С экономической точки зрения, модель показывает, что меры по выводу

экономики из тени имеют затухающую эффективность. Для долгосрочного роста легальных секторов необходимо не только создавать начальные условия с эффектом насыщения, но и стимулировать переход работников в более устойчивые формы, такие как индивидуальное предпринимательство и корпоративный сектор.

Список использованных источников

1. Hart K. (1973). Informal Income Opportunities and Urban Employment in Ghana//The Journal of Modern African Studies, 11(1), 61-89.
2. International Labour Office. (2018). Women and Men in the Informal Economy: A Statistical Picture. Geneva: ILO.
3. Lewis W.A. (1954). Economic Development with Unlimited Supplies of Labour// The Manchester School, 22(2), 139-191.
4. Harris J.R., Todaro M.P. (1970). Migration, Unemployment and Development: A Two-Sector Analysis// American Economic Review, 60(1), 126-142.
5. Bosch M., Esteban-Pretel J. (2012). Job creation and job destruction in the presence of informal markets// Journal of Development Economics, 98(2), 270-286.
6. Meghir C., Narita R., Robin J.-M. (2015). Wages and Informality in Developing Countries// American Economic Review, 105(4), 1509-1546.
7. Esteban-Pretel J., Kitao S. (2021). Labor Market Policies in a Dual Economy// Labour Economics, 68, 101956. <https://doi.org/10.1016/j.labeco.2020.101956>.
8. Loayza N.V. (1996). The economics of the informal sector: a simple model and some empirical evidence from Latin America// Carnegie-Rochester Conference Series on Public Policy, 45, 129-162.
9. Schneider F., Enste D H. (2000). Shadow Economies: Size, Causes, and Consequences// Journal of Economic Literature, 38(1), 77-114.
10. Maloney W.F. (2004). Informality Revisited// World Development, 32(7), 1159-1178.
11. Perry G. et al. (2007). Informality: Exit and Exclusion. - Washington, DC: World Bank.
12. La Porta R., Shleifer A. (2014). Informality and Development// Journal of Economic Perspectives, 28(3), 109-126.
13. Ulyssea G. (2020). Informality: Causes and Consequences for Development// Annual Review of Economics, 12, 525-546.
14. Hairer E., Nørsett S.P., Wanner G. (1993). Solving Ordinary Differential Equations I: Nonstiff Problems. - Berlin: Springer.
15. Butcher J.C. (2003). Numerical Methods for Ordinary Differential Equations. - Chichester: John Wiley & Sons.
16. Lambert J.D. (1991). Numerical Methods for Ordinary Differential Systems: The Initial Value Problem. - Chichester: John Wiley & Sons.
17. Verhulst P.-F. (1838). Notice sur la loi que la population suit dans son accroissement// Correspondance Mathématique et Physique, 10, 113-121.

3. ПРОГРАММНЫЕ И ТЕЛЕКОММУНИКАЦИОННЫЕ СИСТЕМЫ

Беспалов П.В.

УПРАВЛЕНИЕ ГЕТЕРОГЕННЫМИ ДАННЫМИ В БОЛЬШИХ СИСТЕМАХ

Воронежский государственный технический университет

1. Проблемы обработки гетерогенных данных в больших системах

Цифровая трансформация коренным образом трансформировала характер корпоративной деятельности. Предприятия осуществляют перевод в электронную форму не только учетных операций, но и управленческих, производственных и аналитических процессов, причем каждый подобный переход сопровождается лавинообразным накоплением информационных массивов. Параллельно наблюдается множественность инструментов их обработки, а методы хранения данных подвергаются усложнению. Следствием данной динамики становится ситуация, при которой традиционные средства управления данными перестают обеспечивать адекватную реакцию на возросшую нагрузку: разнородность источников и форматов превышает возможности подходов, рассчитанных на однородную информационную среду [1, 2]. Согласно данным аналитической компании Fortune Business Insights, по итогам 2024 года объем глобального рынка инструментов больших данных достиг 348,21 млрд долларов США, а прогнозируемый среднегодовой темп роста составит 13,5%, что к 2032 году обеспечит увеличение рынка до 961,89 млрд долларов США [1]. Стремительный рост рынка непосредственно коррелирует с обострением проблемы гетерогенности данных, что усиливает актуальность разработки специализированных методов их интеграции. Отечественные исследования в области применения больших данных также фиксируют усиление многообразия источников, носителей и форматов информации как одну из ключевых характеристик современных массивов данных, что требует разработки специализированных методов их обработки и интеграции [2]. Существующая потребность в повышении эффективности обработки информации с учетом роста скорости ее генерации и объемов дополняется увеличением ее разрозненности, что формирует запрос на качественно новые подходы к интеграции данных.

Отечественный контекст цифровой трансформации характеризуется наличием специфических особенностей. Российские предприятия в большинстве секторов экономики исторически формировали свои информационные системы фрагментарно: автоматизация начиналась с отдельных участков производства или управления и лишь впоследствии распространялась на смежные процессы, что порождало устойчивые информационные разрывы между подразделениями. В государственном управлении отдельные ведомства на протяжении десятилетий эксплуатировали изолированные реестры и базы данных, разработанные по различным стандартам и с разной степенью структурированности данных. Реализация масштабных проектов цифровизации как в промышленности, так и в государственном секторе все острее ставит вопрос об эффективных методах согласования разнородных источников информации:

необходимость своевременного предоставления полной и непротиворечивой информации для принятия управленческих решений становится ключевым требованием к современным информационным системам [2].

Гетерогенность представляет собой не просто технический недостаток, а фундаментальную проблему, лежащую в основе работы с данными из множества источников. Данные, которым свойственны разрозненность и различия в типах, моделях, языках описания процедур и структурах, принято характеризовать как гетерогенные [9]. Невозможность отказаться от работы с такими данными предъявляет повышенные требования к системам, обеспечивающим их хранение и обработку. Принципиальное отличие гетерогенности от обычной задачи преобразования форматов является более глубоким, чем представляется на первый взгляд. Она затрагивает несовместимость концептуальных моделей, в которых разные системы описывают одни и те же объекты реального мира. Главной проблемой при организации хранения и доступа к информации в крупных распределенных системах является именно неоднородность имеющихся ресурсов: огромные объемы данных хранятся в различных СУБД, однако извлечь из них согласованную картину без специальной предобработки зачастую не представляется возможным [8]. Задача интеграции данных заключается в объединении сведений из разрозненных источников в полный непротиворечивый набор [9].

В интеграции данных принято выделять задачи обеспечения совместимости на уровне значений, синтаксиса, структуры и семантики, каждая из которых предполагает собственные методы реализации [9]. Структурная гетерогенность проявляется в различиях организации схем данных при описании одних и тех же предметных областей. Прежде чем рассматривать примеры, необходимо уточнить терминологию: в контексте реляционных и структурированных данных под полем (или атрибутом, столбцом) понимается именованная характеристика объекта - элемент структуры таблицы или записи, определяемый своим наименованием и типом данных. Совокупность полей образует схему данных. Так, типичным примером структурной гетерогенности является атрибут «адрес»: одна система хранит его единой строкой, другая разбивает на отдельные поля «город», «улица», «дом». Схожая ситуация складывается с именованием атрибутов: поле `client\_id` в одной базе данных соответствует полю `customer\_number` в другой, а поле `full\_name` разделено на `first\_name` и `last\_name` в третьей. Далее термины «поле», «атрибут» и «столбец» используются как синонимы в зависимости от контекста: «атрибут» - при семантическом анализе схем, «столбец» - при описании табличного представления, «поле» - в общем техническом смысле.

Для дальнейшего изложения целесообразно ввести формальное представление схемы данных. Схему источника i определим как пару $S_i = (A_i, T_i)$, где $A_i = \{a_1, \dots, a_n\}$ - множество атрибутов источника, а $T_i: A_i \rightarrow \text{Types}$ - функция, сопоставляющая каждому атрибуту его тип данных (числовой, категориальный, строковый, дата и т. п.). Тогда задача сопоставления схем двух источников формально сводится к построению отображения $M: A_1 \times A_2 \rightarrow [0, 1]$, ставящего в соответствие каждой паре атрибутов ($a \in A_1, b \in A_2$) оценку

их семантической эквивалентности $\text{sim}(a, b)$, причём пара считается соответствующей при $\text{sim}(a, b) \geq \theta$, где θ - заданное пороговое значение. Такая формализация позволяет единообразно описать все три ранее рассмотренных вида гетерогенности: структурная выражается несовпадением множеств A_1 и A_2 при семантической близости их элементов, семантическая - высоким значением $\text{sim}(a, b)$ при формальном различии типов $T_1(a) \neq T_2(b)$ или наименований атрибутов, а синтаксическая - различием в способе кодирования значений одного и того же домена.

Семантическая гетерогенность возникает в ситуации, когда идентичные значения имеют различную смысловую интерпретацию в разных системах. Так, числовой код «1» может означать «активный клиент» в CRM-системе и «заблокированный» в биллинговой системе того же предприятия. Более сложные случаи связаны с различием бизнес-правил: понятие «выручка» в одной системе включает налог на добавленную стоимость, в другой - нет, «дата заказа» в одной системе фиксирует момент создания заявки, в другой - момент подтверждения. Такие конфликты не выявляются на уровне схемы и требуют глубокого семантического анализа содержимого данных, что существенно усложняет задачу автоматической интеграции.

Синтаксическая гетерогенность выражается в различиях форматов представления данных. Разные системы используют несовпадающие форматы дат ('DD.MM.YYYY', 'YYYY-MM-DD'), кодировки строк (UTF-8, Windows-1251), разделители в числовых значениях, различные булевы обозначения ('true'/'false', '1'/'0', 'Y'/'N'). Хотя синтаксические конфликты технически проще остальных, отсутствие автоматизации их устранения требует значительных трудозатрат. Анализ алгоритмов сопоставления схем данных показывает, что современные инструменты вынуждены учитывать все три вида гетерогенности одновременно, применяя комбинированный структурный, семантический и атрибутивный анализ, средняя эффективность существующих решений не превышает 26%, а в лучших случаях достигает 40–60% верно определенных связей между атрибутами [4].

Источники возникновения гетерогенности данных в корпоративной среде не случайны - они накапливались десятилетиями. По мере роста бизнеса и расширения продуктовой линейки появляются новые подразделения, каждое из которых формирует собственные структуры хранения данных, ориентированные на специфику своих задач. Разные отделы компании работают с различными задачами, из-за чего скорость, качество и объем поступающих данных существенно различаются, а хранение информации может осуществляться как централизованно в корпоративном хранилище, так и на уровне отдельных подразделений. Последнее повышает риск дублирования данных, а их актуализация разными сотрудниками нередко приводит к противоречиям [9]. Источники данных, как правило, не ограничиваются пределами одной организации: информация поступает от партнеров, клиентов и контрагентов с принципиально различными технологическими решениями и специфическими форматами выгрузок. Наконец, существенным фактором выступает проблема унаследованных систем. Базы данных зачастую создаются фрагмен-

тарно для решения конкретных задач, однако по прошествии времени их функционал расширяется, структуры подвергаются регулярному пересмотру, а нарастание числа несогласованных структур данных влечет за собой увеличение количества проблем, связанных с организацией данных [8].

Каждый класс систем представляет собой изолированный сегмент. Многообразие классов информационных систем, функционирующих в пределах одного предприятия, является ключевым фактором, усугубляющим проблему гетерогенности. Системы управления ресурсами предприятия (ERP) оперируют финансовыми транзакциями, плановыми показателями и нормативно-справочной информацией в структурированном реляционном формате. Системы управления взаимоотношениями с клиентами (CRM) накапливают полуструктурированные сведения о взаимодействиях, историю коммуникаций и поведенческие данные. Системы управления производством (MES) собирают телеметрию с оборудования и данные о технологических процессах в режиме реального времени. Системы бизнес-аналитики (BI) агрегируют показатели из множества первичных систем для построения управленческой отчетности. Каждый из перечисленных классов формирует собственную логическую модель данных, собственные идентификаторы сущностей и собственные правила валидации. При необходимости кросс-функционального анализа - оценки влияния производственных сбоев на выполнение клиентских заказов - все эти системы должны быть интегрированы. Без специализированных инструментов согласования это требует значительных ручных затрат (рис. 1).



Рис. 1. Классификация информационных систем

Необходимость выстраивания процессов преобразования и форматирования данных при переходе на новый инструментарий остается одной из ключевых сложностей в работе с гетерогенными источниками. Зачастую бывает затруднительно использовать информацию, накопленную в устаревших системах, из-за проблем с совместимостью форматов. Такие исторические данные остаются ценным ресурсом для организаций, однако трудности при попытке их извлечения и интеграции делают этот ресурс труднодоступным [8]. Особую сложность представляет проблема дублирования и избыточности: схожие данные из разных источников могут описывать один и тот же объект - клиента, сотрудника, продукт - посредством различных идентификаторов и атрибутов, что делает невозможным их автоматическое объединение без специализированных алгоритмов сопоставления. Следствием становится снижение качества аналитики, ошибки в отчетности и дополнительные трудозатраты на ручное согласование данных, которые в условиях больших систем становятся неприемлемо высокими.

Задача интеграции гетерогенных данных в крупных системах традиционно решается с помощью процессов извлечения, преобразования и загрузки данных - ETL (Extract, Transform, Load). ETL предоставляет инструментарий для извлечения данных из разрозненных источников, их преобразования в единый формат и загрузки в целевое хранилище [3]. Критическим недостатком физической интеграции данных, реализуемой через ETL, является ее негибкость. Несмотря на высокую производительность обработки запросов, подход требует значительных затрат на проектирование трансформационных правил, причем при любом изменении схемы источника необходима полная переработка всего конвейера, что делает его непригодным для динамичных сред. Кроме того, ETL реализует пакетную обработку данных по расписанию, что исключает его применение для задач, требующих актуальности данных в реальном времени [3]. Виртуальная интеграция, при которой источники запрашиваются напрямую через промежуточный слой без физического копирования, служит более гибкой альтернативой, однако существенно усложняется при наличии множественных конфликтов гетерогенности: каждый новый источник требует разработки специфических адаптеров и трансляторов. В обоих случаях отсутствие единого языка описания схем источников вынуждает интеграционные системы опираться на специализированные запросы к каждому отдельному источнику [3], что принципиально ограничивает масштабируемость при работе с десятками и сотнями разнородных источников. Это ограничение наглядно представлено на рис. 2.

Осознание ограничений ETL-подходов стимулировало развитие методов семантической интеграции данных - подхода, при котором акцент смещается с технического преобразования форматов на согласование смыслового содержания данных. Семантическая интеграция предполагает разработку единого контекста, в котором разрозненные данные из различных источников получают согласованную интерпретацию.

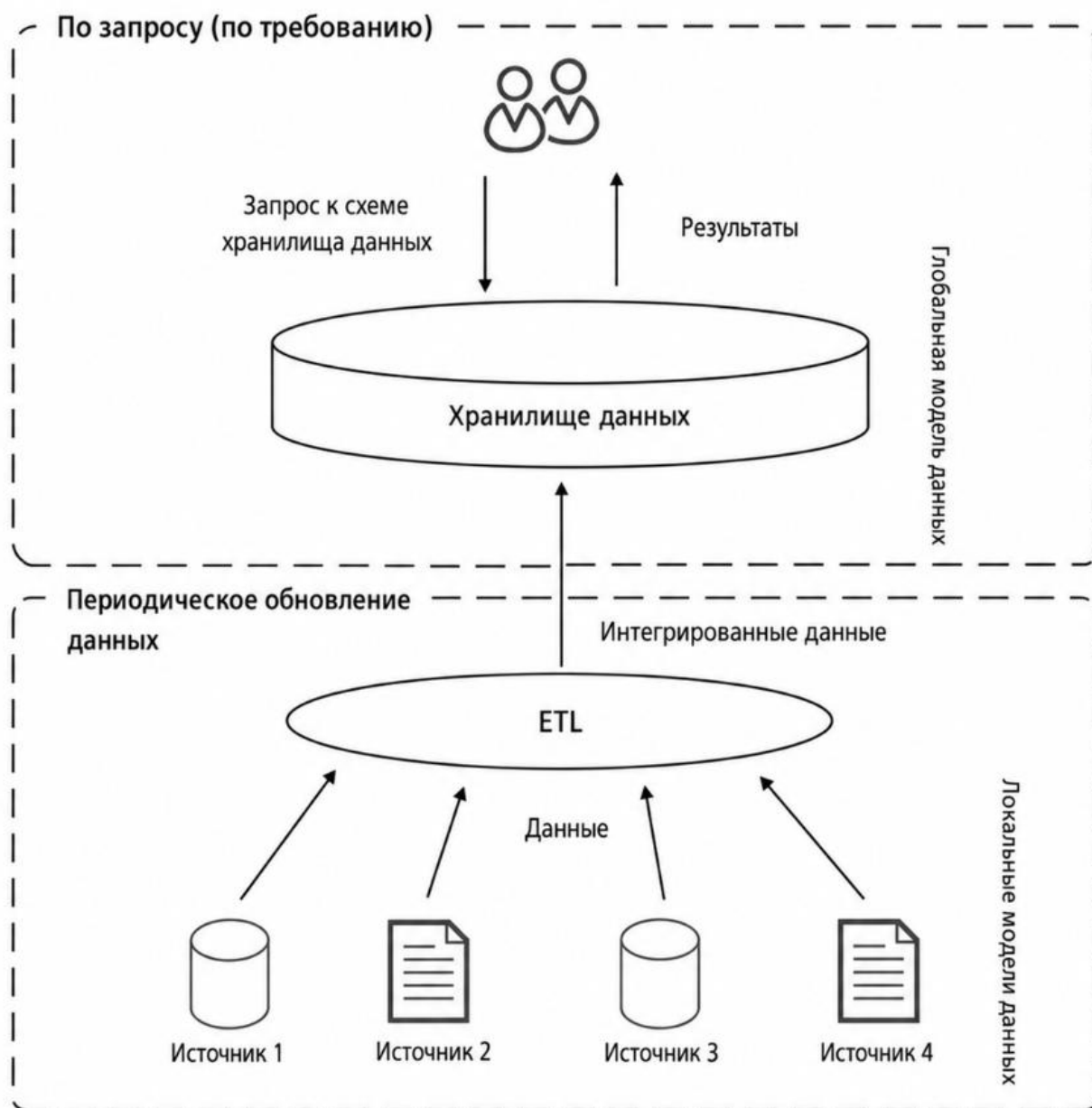


Рис. 2. Архитектура хранилища данных на основе ETL [3]

Ключевым инструментом семантической интеграции является онтология - формализованное описание предметной области, включающее словарь понятий, их свойства и семантические связи между ними [9]. Однако построение и поддержание онтологий требует значительных экспертных усилий, что ограничивает их применимость в динамичных корпоративных средах. Наличие единой онтологии позволяет сопоставлять термины из разных систем через их смысловые связи даже при различии наименований: так, «контрагент» в одной системе и «поставщик» в другой могут быть однозначно соотнесены через иерархию онтологических понятий. Задача маппинга данных, установление соответствий между атрибутами различных источников в онтологическом подходе, решается посредством поиска семантических связей между элементами разных онтологий. Постоянное обновление и актуализация используемого словаря метаданных обеспечивает устойчивость такого подхода в условиях динамично меняющейся бизнес-среды [9]. Тем не менее семанти-

ческая интеграция на основе онтологий также сопряжена с существенными трудозатратами при построении и поддержании онтологий, что обуславливает необходимость их автоматизации средствами машинного обучения и алгоритмического согласования схем. Подход на основе онтологий изображен на рис. 3.



Рис. 3. Классификация методик сопоставления сущностей [4]

Проблема обработки гетерогенных данных в больших системах является комплексной и многоаспектной. Ее природа вытекает из объективных закономерностей развития информационных систем: независимой разработкой источников, технологической эволюцией предприятий и организационной динамикой. Структурная, семантическая и синтаксическая гетерогенность совместно формируют барьеры для автоматической интеграции, преодоление которых средствами традиционных ETL-подходов ограничено как по масштабируемости, так и по актуальности данных. Ни один из рассмотренных подходов - ни физическая ETL-интеграция, ни виртуальная, ни онтологическая - не решает задачу автоматического согласования схем и кластеризации записей без участия эксперта или размеченных данных. Этот методологический разрыв между существующими решениями и практическими потребностями обосновывает разработку специализированных алгоритмических подходов, способных работать в условиях отсутствия априорной разметки, что и составляет предмет дальнейшего исследования.

2. Методы согласования схем данных и алгоритмы кластеризации

Преодоление барьеров структурной, семантической и синтаксической гетерогенности требует специализированных алгоритмических средств, способных автоматически выявлять соответствия между источниками и группировать связанные записи. Интеграция разнородных данных из множества независимых источников является многоуровневым процессом, каждый этап которого требует специализированных методов. Вовченко А.Е., Калиниченко Л.А. и Ковалев Д.Ю. формализуют общую структуру этого процесса в виде последовательного конвейера, включающего пять этапов: сопоставление схем (Schema Matching), интеграцию схем (Schema Integration), трансформа-

цию данных (Data Transformation), разрешение сущностей (Entity Resolution) и слияние данных (Data Fusion) [10]. Эта декомпозиция существенна, поскольку показывает, что задача согласования данных распадается на два качественно различных уровня - уровень схем, где согласуется структура источников, и уровень записей, где отождествляются конкретные объекты реального мира. Оба уровня требуют самостоятельных алгоритмических решений, причем качество работы на уровне записей напрямую определяется полнотой и точностью сопоставления схем. Отсутствие единой схемы данных и семантической согласованности между источниками делает каждый этап конвейера нетривиальной алгоритмической задачей [10].

На уровне схем задача сводится к сопоставлению схем данных - процедуре выявления семантических и структурных взаимосвязей между атрибутами различных источников с целью установления их соответствия. Кургузкин К.Н. и Марьенков А.Н. систематизируют накопленный опыт в данной области, разграничивая два принципиально различных класса методик [4]. Экспертные методики опираются на ручное определение соответствий специалистами предметной области; при высокой точности они плохо масштабируются и требуют значительных трудозатрат при большом числе атрибутов или источников. Алгоритмические методики автоматизируют этот процесс и в зависимости от используемого типа анализа делятся на три направления. Синтаксическое - сопоставление по именам, типам данных, форматам, регулярным выражениям и синтаксическим N-граммам, разбивающим имена атрибутов и их значения на токены для последующего сравнения. Семантическое - использование словарей, тезаурусов, онтологий (в частности, WordNet) и методов латентного семантического анализа для выявления смысловых соответствий между атрибутами с различными именами, но одной природой (например, «контактный телефон» и «мобильный телефон»). Структурное - анализ иерархий и связей между атрибутами, при котором два атрибута признаются соответствующими, если занимают аналогичное место в структурах своих схем [4].

Ни один из этих типов анализа по отдельности не дает приемлемой точности в реальных условиях. Синтаксический метод не распознает смысловые синонимы - `client\_id` и `customer\_number` останутся несопоставленными. Семантический анализ через WordNet игнорирует контекст расположения атрибута в схеме. Структурный подход теряет эффективность при различии топологий источников - две схемы могут описывать один объект через разное число таблиц и связей.

Одним из способов преодоления этих ограничений служит графовый подход, в котором схема представляется ориентированным графом: вершины соответствуют атрибутам, ребра отражают их взаимосвязи [4]. Тем не менее при значительном различии структур двух источников графовое сопоставление может выдать ложные соответствия между атрибутами, имеющими схожее окружение, но разный смысл. Это обуславливает необходимость комбинирования нескольких типов анализа с адаптивным взвешиванием их вкладов. Общая структура процесса сопоставления схем данных, включающая

этапы анализа источников, подготовки данных, непосредственного сопоставления атрибутов и формирования решения по интеграции, представлена на рис. 4.

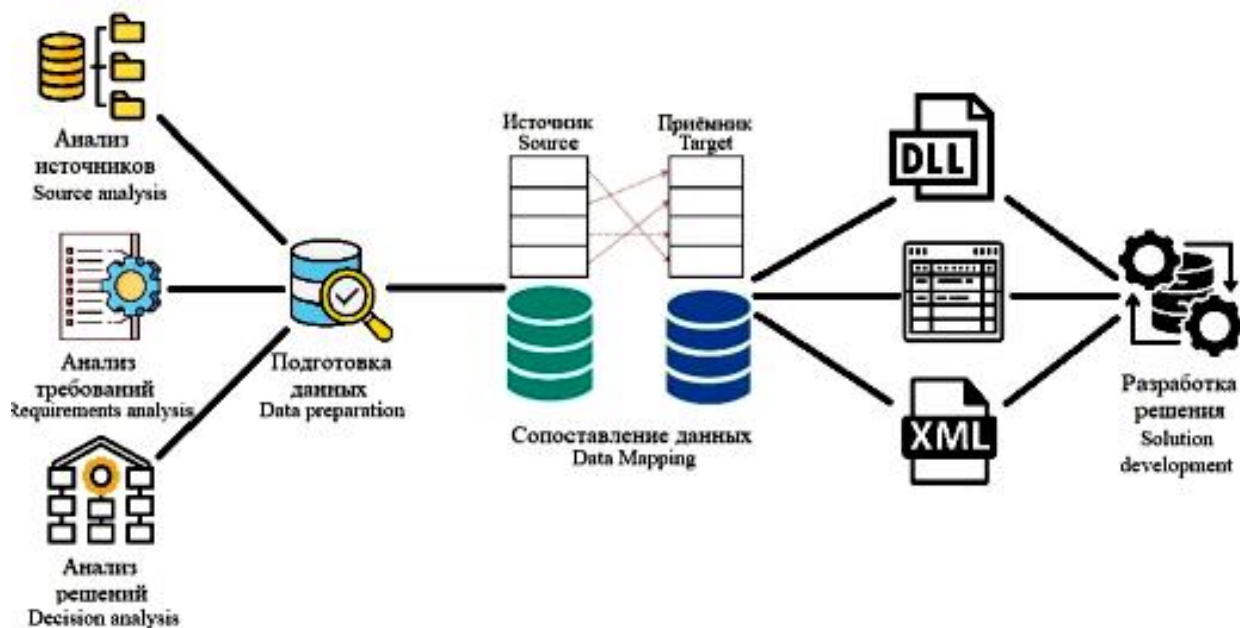


Рис. 4. Графическая аннотация процесса сопоставления схем данных

Экспериментальное сравнение трех распространенных инструментов («OpenII», «Coma-3.0» и «Rondo») на разнообразных тестовых сценариях выявило существенный разрыв между теоретическими возможностями методов и их практической эффективностью: средняя результативность по всем сценариям не превысила 26% верно определенных соответствий, а в лучших отдельных случаях - 40–60% [4]. Причина проста. Каждый из инструментов реализует преимущественно один тип анализа и не обеспечивает их совместного применения. Этот вывод обосновывает разработку гибридного алгоритма сопоставления, объединяющего несколько типов анализа с адаптивным взвешиванием вкладов каждого из них.

Примером такой архитектуры служит система Harmony [4]: она запускает несколько метчеров параллельно (структурный, семантический и синтаксический) и объединяет их оценки через взвешенное среднее, что позволяет каждому из компонентов компенсировать слабости остальных. Именно независимые компоненты анализа, объединяемые через взвешенное среднее с настраиваемыми весами, рассматриваются в литературе как наиболее перспективный подход к построению гибридных систем сопоставления схем. Формально итоговая оценка сходства пары атрибутов в подобных гибридных системах вычисляется по формуле (1):

$$\text{sim}(a, b) = \sum_{k=1}^K w_k \cdot \text{sim}_k(a, b), \quad \sum w_k = 1, w_k \geq 0 \quad (1)$$

где $\text{sim}_k(a, b)$ - оценка сходства, полученная k -м метчером (структурным, семантическим или синтаксическим), а w_k - вес, отражающий относительную надёжность метчера k для конкретной пары источников. Настройка весов w_k вручную либо посредством обучения на размеченной выборке соответствий составляет основную сложность практической реализации гибридного под-

хода. Но у самой архитектуры Harmony есть существенное слабое место. Авторы Кургузкин К.Н. и Марьенков А.Н. опираются на англоязычные тезаурусы (WordNet - лексическую базу данных, группирующую слова по смысловым отношениям) и не рассматривают случай, когда семантический компонент должен работать с русскоязычными метаданными. Прямая переносимость их выводов ограничена: словарный семантический анализ приходится заменять эмбединг-моделью, поддерживающей русский язык. Сравнительная характеристика рассмотренных классов методов сопоставления схем по ключевым практическим параметрам представлена в табл. 1.

Таблица 1

Сравнение классов методов сопоставления схем данных по ключевым практическим параметрам

Класс метода	Пример	Точность	Масштабируемость	Ручные затраты
Экспертный	Ручное сопоставление	Высокая	Низкая	Высокие
Синтаксический	N-граммы, типы данных	Низкая	Высокая	Низкие
Семантический	WordNet, LSA	Средняя	Средняя	Средние
Структурный	Изоморфизм графов	Средняя	Низкая	Низкие
Гибридный	Harmony	Высокая	Средняя	Низкие

Если задача сопоставления схем относится к уровню структуры источников, то следующий этап интеграции переносит фокус на уровень содержания - конкретные записи. На этом этапе центральной становится задача разрешения сущностей (entity resolution) - отождествление записей из разных источников, описывающих один и тот же объект реального мира. Вовченко А.Е. и соавторы определяют разрешение сущностей как извлечение информации об одной и той же сущности из разнообразных структурированных коллекций данных и приведение ее к унифицированному представлению [10]. Под этим термином объединяется несколько взаимосвязанных подзадач: выявление дубликатов, дедупликация, связывание записей, идентификация объектов и сопоставление ссылок [10].

Все перечисленные задачи объединяет одна вычислительная сложность: при попарном сравнении N записей число операций составляет $O(N^2)$, что при миллионных объемах данных делает наивный подход неприменимым. Существенное отличие разрешения сущностей от однократного попарного сравнения (data matching) заключается в следующем. В отличие от последнего, разрешение сущностей работает итеративно: записи постоянно пополняются данными из новых источников, а связи между ними уточняются по мере поступления информации, что позволяет постепенно приближаться к наиболее точному и полному представлению каждой сущности. Процесс разрешения сущностей включает нормализацию данных, вычисление мер сходства для каждого типа атрибутов и применение метода разрешения пар с учетом логических ограничений, транзитивности и эксклюзивности [10].

Последний этап конвейера интеграции - слияние данных (data fusion) -

задача объединения согласованных сущностей в единое непротиворечивое представление с разрешением возможных конфликтов между источниками. Различные аспекты этой задачи (типы конфликтов, стратегии их разрешения, операторы объединения и функции вычисления итоговых значений) систематизированы в работе [10] и отражены на рис. 5. Хотя слияние данных не является центральной задачей разрабатываемого программного обеспечения, понимание его места в общей структуре конвейера необходимо для корректной постановки задачи разрешения сущностей: точность последнего непосредственно определяет качество последующего слияния.

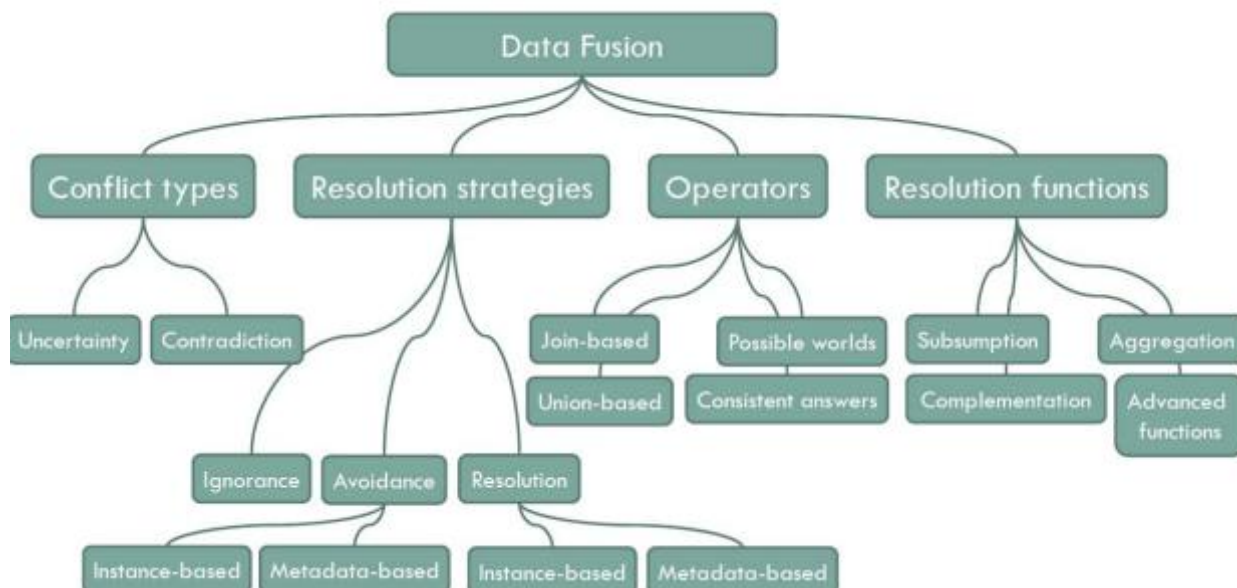


Рис. 5. Различные аспекты проблемы слияния данных [10]

Дедупликация сводится к удалению точных или близких дубликатов из одного массива данных. Плотникова Н.П., Кевбрин В.А. и Болохов Д.А. исследуют эту задачу применительно к большим объемам неструктурированных данных, подчеркивая, что избыточность не только расходует ресурсы хранения, но и систематически искажает результаты обучения нейронных сетей на дублирующихся примерах [5]. Авторы анализируют три подхода с позиций применимости к большим данным. Подход на основе множеств в памяти ограничен объемом RAM и неприменим при объемах, превышающих его. Построчное чтение с внешним файлом-реестром не ограничено памятью, однако страдает от квадратичной сложности поиска. Подход с использованием СУБД устраняет узкое место поиска за счет индексирования: хранение хэш-значений записей, вычисленных по алгоритму MD5 (Message Digest 5, криптографическая хэш-функция, формирующая 128-битный код фиксированной длины), вместо самих данных обеспечивает независимость от типа данных и минимальные затраты памяти, а ограничение UNIQUE (ограничение целостности SQL, гарантирующее уникальность значений в столбце) на уровне схемы базы гарантирует структурное отсутствие дублей [5]. Но здесь кроется принципиальное ограничение. Все три подхода рассчитаны на точную дедупликацию по полному совпадению хэшей и не работают там, где записи описывают один объект с различиями в написании, формате и полно-

те атрибутов.

Однако описанный подход имеет существенное ограничение: хэш-сравнение идентифицирует лишь точные дубликаты - записи, побайтово совпадающие после нормализации. Записи с опечатками, сокращенными именами, иным порядком слов или разными форматами дат хэш-подход не обнаруживает. Это обуславливает необходимость применения нечетких метрик сходства на уровне отдельных атрибутов, что переводит задачу в область вычисления функций расстояния. Однако, прежде чем переходить к самим метрикам, необходимо определить, как покомпонентные оценки сходства объединяются в единую меру для пары записей. Именно стратегия агрегирования определяет, как метрики на уровне отдельных полей превращаются в решение о соответствии записей в целом.

Выбор метода агрегирования оценок сходства по отдельным атрибутам в единую меру для пары записей определяет всю стратегию разрешения сущностей. Вовченко А.Е. и соавторы рассматривают три основных подхода [10]. Детерминированный взвешенный подход формирует итоговую меру как линейную комбинацию оценок по атрибутам. Его слабость - необходимость ручного подбора весов и порогового значения, что превращается в нетривиальную задачу при нескольких десятках атрибутов. Подход на основе булевых правил задает предикаты по атрибутам; при большей интерпретируемости он еще менее масштабируем - каждое новое сочетание атрибутов требует ручного правила. Вероятностный подход на основе модели Феллеги-Сантера снимает необходимость ручного задания весов: он рассматривает пары записей как выборку из двух распределений (совпадающих и несовпадающих) и классифицирует каждую пару по отношению правдоподобия, автоматически оценивая веса из данных [10]. Формально решение о статусе пары записей в этой модели принимается на основе отношения правдоподобия (2):

$$R(\gamma) = P(\gamma | M) / P(\gamma | U) \quad (2)$$

где γ - вектор совпадений по отдельным атрибутам пары записей, а M и U - гипотезы «совпадение» и «несовпадение» соответственно; пара классифицируется как совпадение при $R(\gamma) \geq T_{upper}$, как несовпадение при $R(\gamma) \leq T_{lower}$ и как требующая ручной проверки в промежуточной зоне [10]. Однако на практике метод требует размеченной выборки для оценки вероятностей $P(\gamma | M)$ и $P(\gamma | U)$, что без априорной разметки источников ограничивает его применение.

Вместе с тем все три подхода объединяет общее ограничение: ни один из них не учитывает специфику типа атрибута при выборе функции сходства - веса назначаются атрибутам одинаково, независимо от того, является ли поле числовым, категориальным или текстовым. Именно этот пробел восполняет адаптивная функция расстояния, которая дифференцирует метрики по типу атрибута. Формально её можно определить как кусочно заданную функцию (3):

$$d(r_i, r_j) = \sum_{k=1}^m w_k \cdot \delta_k(r_i^k, r_j^k) / \sum_{k=1}^m w_k \quad (3)$$

где m - число сопоставленных атрибутов пары записей r_i и r_j , w_k - вес атрибута k , а δ_k - покомпонентная функция расстояния, выбираемая в зависимости

от типа атрибута k : расстояние Гауэра - для числовых и категориальных полей (формула 4), одна из строковых метрик (LCS, Жаро-Винклера, Дамерау-Левенштейна - табл. 2) - для строковых полей, причём конкретная строковая метрика выбирается по характеру поля: Жаро-Винклера - для коротких полей (имена, коды), LCS - для полей с вариативным написанием, Жаккарда - для многословных текстовых полей. Такое построение функции d принципиально отличает предлагаемый подход от рассмотренных выше методов агрегирования, ни один из которых не учитывал тип атрибута при выборе метрики.

Методы машинного обучения (деревья решений, SVM, ансамбли классификаторов и условные случайные поля CRF) переводят задачу формулировки критериев сопоставления из ручного режима в режим обучения по данным. Однако они порождают новый тип трудностей: несбалансированность классов (несхожих пар на несколько порядков больше, чем схожих) систематически смещает модель к предсказанию «несовпадение», а ключевая проблема - необходимость формирования размеченной обучающей выборки - особенно остро встает при интеграции разнородных данных без априорной разметки. Если детерминированные методы страдают от трудоемкости ручного задания правил, то supervised-методы МО переносят эту трудоемкость в область разметки данных, что при отсутствии готового корпуса делает их применение затруднительным. Методы без учителя и активного обучения, хотя и перспективны теоретически, на практике требуют значительных исследовательских усилий для адаптации к специфике гетерогенных данных [10]. Поэтому при неизвестной структуре источников в настоящей работе принят детерминированный подход с адаптивно вычисляемой функцией расстояния: он не требует предварительной разметки и обеспечивает интерпретируемость результатов, что критично для корпоративных приложений.

Независимо от выбранного метода агрегирования, центральным вопросом остается измерение сходства отдельных атрибутов. Принципиальная трудность состоит в разнотипности атрибутов: числовые, категориальные и строковые поля требуют различных функций расстояния, а их прямое смешивание без нормировки приведет к доминированию атрибутов с большим диапазоном значений. Классическим решением этой проблемы служит расстояние Гауэра - обобщенная метрика, допускающая совместную работу с переменными разных типов и шкал [8].

Формально расстояние Гауэра определяется как взвешенная сумма покомпонентных вкладов по формуле (4):

$$S_{ij} = \frac{\sum_{k=1}^p w_{ijk} s_{ijk}}{\sum_{k=1}^p w_{ijk}} \quad (4)$$

где s_{ijk} - вклад признака k в сходство объектов i и j , а w_{ijk} - весовой коэффициент, равный 0 при пропущенном значении и 1 иначе. Для количественных признаков $s_{ijk} = 1 - |x_{ik} - x_{jk}| / (R - k)$ - диапазон значений признака по всей выборке; для качественных - при совпадении категорий и 0 иначе. Нормировка обес-

печивает сопоставимость атрибутов с разными единицами измерения [8].

Однако для текстовых атрибутов метрика Гауэра неприменима: она использует лишь бинарное совпадение/несовпадение, тогда как в реальных данных имена и адреса часто содержат опечатки и перестановки. Для работы с текстовыми атрибутами необходим отдельный класс метрик. Рубцов Д.Н. и Барахнин В.Б. проводят сравнительный анализ строковых метрик подобия применительно к дедупликации записей в разнородных библиографических базах данных [9].

Расстояние Левенштейна определяется как минимальное число операций вставки, удаления и замены символов для преобразования одной строки в другую; вычисляется методом динамического программирования со сложностью $O(|s^1| * |s^2|)$ [9]. Расстояние Жаро-Винклера ориентировано на короткие строки с опечатками: поправка Винклера дополняет базовую метрику учётом длины общего префикса, что повышает точность при сопоставлении имён и фамилий, для которых характерны совпадения первых символов [9]. Расстояние Дамерау-Левенштейна обобщает расстояние Левенштейна за счёт добавления четвёртой операции - перестановки двух соседних символов, что позволяет точнее моделировать типичные опечатки при ручном вводе; экспериментально установлено, что при длине строки от 5 до 9 символов точность классификации превышает 0,9, а при длине 20-30 символов достигает 0,97 [11]. Коэффициент Жаккарда для строк сравнивает множества биграмм и вычисляется как отношение мощности их пересечения к мощности объединения, что обеспечивает устойчивость к перестановкам слов в многословных полях [9]. Тестирование метрик на реальных базах данных установило, что метрика на основе наибольшей общей подпоследовательности (LCS) обеспечивает наилучшую эффективность при поиске дубликатов в гетерогенных источниках, тогда как Жаро-Винклера предпочтительна для коротких полей с единичными опечатками [9]. Специализированные строковые метрики решают проблему частичного сходства строк, однако каждая ориентирована на определённый тип искажений (табл. 2).

Строковые метрики и адаптивные веса позволяют построить матрицу попарных расстояний между всеми записями, однако сама по себе она не даёт ответа на вопрос, какие записи описывают одну сущность. Для этого требуется алгоритм, выделяющий в матрице плотные группы взаимно близких объектов. Выбор алгоритма кластеризации не менее критичен, чем выбор функции расстояния. Жилов Р.А. систематизирует основные методы кластеризации, разграничивая иерархические алгоритмы и неиерархические, заранее требующие задания числа групп [6].

Среди неиерархических метод k-средних отличается простотой и высокой скоростью, однако принципиально непригоден для задач интеграции данных: он требует задания числа кластеров k заранее, что при неизвестном количестве уникальных сущностей невозможно; чувствителен к начальной инициализации; неспособен обнаруживать кластеры произвольной формы; наконец, включает выбросы и ошибочные записи в ближайший кластер, искажая результат [6].

Таблица 2

Сравнение строковых метрик подобия по ключевым параметрам применимости

Метрика	Учитываемые операции	Сложность	Опечатки	Перестановки	Оптимально для
Жаро-Винклера	Совпадение, транспозиция, общий префикс	$O(mn)$	Да	Частично	Коротких имён и фамилий
Левенштейна	Вставка, удаление, замена	$O(mn)$	Да	Нет	Общих строковых полей
Хэмминга	Только замена	$O(n)$	Нет	Нет	Строк одинаковой длины
Дамерау-Левенштейна	+ Транспозиция соседних	$O(mn)$	Да	Да	Коротких полей с перестановками
LCS	Наибольшая общая подпоследовательность	$O(mn)$	Частично	Нет	Полей с вариативным написанием
Жаккарда	Множества биграмм	$O(n)$	Нет	Да	Многословных текстовых полей

Нейросетевые методы (SOM и сети Кохонена) способны выявлять нелинейные зависимости и визуализировать кластерную структуру в двумерном пространстве, однако также требуют предварительного задания числа нейронов и вычислительно затратны [6]. Это делает обе группы неподходящими для задачи согласования разнородных данных, где число уникальных сущностей заранее неизвестно, а присутствие шума - типичная ситуация. Критерий заблаговременного задания k является определяющим, при интеграции реальных источников количество уникальных сущностей варьируется от десятков до тысяч и может быть установлено только по результатам кластеризации, но не до неё.

Пестунов И.А. и Синявский Ю.Н. в обширном сравнительном обзоре формулируют пять требований к алгоритму, пригодному для работы с большими разнородными данными: низкая вычислительная сложность; способность выделять кластеры разной формы, размера и плотности при минимальных априорных знаниях; определение числа кластеров из данных без его предварительного задания; устойчивость к шуму и выбросам; простота настройки параметров [7].

Проверяя шесть групп алгоритмов (разбивающие, иерархические, плотностные, сеточные, нейросетевые и комбинированные) по каждому из критериев, авторы устанавливают, что только плотностные методы (прежде всего DBSCAN и его расширения) удовлетворяют одновременно требованиям автоматического определения числа кластеров и устойчивости к шуму [7]. Все прочие группы удовлетворяют не более чем трём из пяти требований: разбивающие требуют задания k и не устойчивы к шуму; иерархические требуют

задания k и плохо масштабируются на больших объёмах данных; нейросетевые требуют задания числа нейронов и вычислительно затратны [7]. На базовых данных без аномалий дендрограммы дивизионной и агломеративной кластеризации образуют относительно компактные ветви с выраженными уровнями объединения, что позволяет визуально выделять групповую структуру. При наличии же аномальных наблюдений структура дендрограмм существенно нарушается, появляются длинные изолированные ветви на крайних позициях, нарушающие монотонность уровней объединения, что подтверждает непригодность иерархических методов для задач интеграции зашумлённых данных [8].

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) формально строится на трёх понятиях: ϵ -соседство, ядерные точки и достижимость (формула 5). Кластер определяется как максимальное связное множество взаимно достижимых точек; все прочие точки классифицируются как шум [7]:

$$N_{\epsilon}(x) = \{y \mid X \mid D(x, y) \leq \epsilon\} \quad (5)$$

Именно понятие шума делает DBSCAN принципиально отличным от k -средних и иерархических методов: зашумлённые записи исключаются из результата, а не включаются в ближайший кластер. В задаче согласования гетерогенных данных это критично - источники часто содержат ошибочные или неполные записи без пар в других источниках.

У DBSCAN есть существенные ограничения, релевантные для практической реализации. Квадратичная сложность при наивной реализации делает алгоритм неприменимым для больших объёмов без предвычисления матрицы расстояний. Единственная пара параметров (ϵ , minPts) не подходит для кластеров с сильно различающейся плотностью. Высокая размерность признакового пространства снижает эффективность [6]. Расширение OPTICS решает проблему неоднородной плотности, но работает в 1,6 раза медленнее [7]. Сравнительная характеристика рассмотренных алгоритмов по ключевым критериям применимости к задаче согласования данных представлена в табл. 3.

Таблица 3

Сравнение алгоритмов кластеризации по критериям применимости к задаче согласования данных

Алгоритм	Требуется задание k	Устойчивость к шуму	Кластеры произвольной формы
k -средних	Да	Нет	Нет
Ward (иерарх.)	Да	Нет	Ограниченно
DBSCAN	Нет	Да	Да
OPTICS	Нет	Да	Да
SOM / Кохонен	Да	Нет	Да

Ни один из рассмотренных методов не закрывает задачу целиком. Инструменты сопоставления схем при однотипном анализе дают 26-60% точности [4]. На уровне записей картина не лучше - детерминированные методы требуют ручного задания правил, методы машинного обучения упираются в отсутствие размеченной выборки [10], а хэш-дедупликация находит лишь точ-

ные совпадения [5]. Универсальные метрики расстояния игнорируют специфику текстовых полей [9]. Из всех алгоритмов кластеризации лишь плотностные сочетают устойчивость к шуму с автоматическим определением числа кластеров [7]. Каждое из этих ограничений по отдельности обходимо, но их совокупность не покрывается ни одним существующим инструментом - и именно она формирует контуры разрабатываемого решения: гибридное сопоставление схем (семантика плюс структура), адаптивная функция расстояния с метриками по типу атрибута и DBSCAN на предвычисленной матрице. При этом система должна работать без размеченных данных и почти без настройки экспертом.

Научная новизна предлагаемого подхода состоит не в отдельных его компонентах - каждый из них по отдельности описан в проанализированной литературе, - а в их сочетании применительно к задаче интеграции русскоязычных корпоративных данных без априорной разметки, которое ранее не рассматривалось. Во-первых, семантический компонент гибридного метчера (формула 1) строится на эмбединг-модели, обученной на русском языке, а не на англоязычном тезаурусе WordNet, как в системе Harmony [4]; тем самым снимается ограничение переносимости, отмеченное выше применительно к работе [4]. Во-вторых, вместо единой метрики расстояния - либо метрики Гауэра, либо одной из строковых метрик (табл. 2) - вводится адаптивная функция расстояния (3), автоматически выбирающая тип метрики по типу атрибута; такое построение функции d не встречается ни у Кисляков А.Н. и Поляков С.В. [8], ни у Рубцова Д.Н. и Барахнина В.Б. [9], рассматривающих метрики Гауэра и строковые метрики отдельно. В-третьих, применение DBSCAN на предварительно вычисленной матрице попарных адаптивных расстояний к задаче разрешения сущностей в гетерогенных корпоративных данных отличается от известного применения плотностных алгоритмов к сегментации спутниковых изображений [7] и к поиску аномалий на основе иерархических методов [8]: контекст (записи с разнотипными атрибутами, а не пространственные точки) и функция расстояния в известных работах не рассматривались совместно. Совокупность этих трёх решений и составляет предмет дальнейшего исследования - разработку и экспериментальную проверку гибридного алгоритма согласования гетерогенных данных, работающего без экспертной разметки.

Список использованных источников

1. [https://www.tadviser.ru/index.php/Статья:Большие_данные_\(Big_Data\)_в_России](https://www.tadviser.ru/index.php/Статья:Большие_данные_(Big_Data)_в_России).
 2. Суринов А.Е. Официальная статистика в условиях цифровой трансформации // Вопросы статистики. - 2023. - Т. 30. - № 1. - С. 5-20.
 3. Marek R. Understanding ETL// ETL with Python. - Apress, 2014. - P. 1-14.
 4. Кургузкин К.Н., Марьенков А.Н. Сравнительный анализ методик сопоставления информации в контексте интеграции схем данных// Прикаспийский журнал. - 2024. - № 2 (66). - С. 16-30.
 5. Плотникова Н.П., Кевбрин В.А., Болохов Д.А. Дедупликация больших объемов данных при помощи баз данных// Инженерный вестник Дона. - 2023. - №9.
 6. Жилов Р.А. Интеллектуальные методы кластеризации данных // Известия Кабардино-Балкарского научного центра РАН. - 2023. - № 6(116). - С. 152-159.
-

7. Пестунов И.А., Синявский Ю.Н. Алгоритмы кластеризации в задачах сегментации спутниковых изображений// Вестник КемГУ. - 2012. - № 4(52). - Т. 2. - С. 109-121.
8. Кисляков А.Н., Поляков С.В. Иерархические методы кластеризации в задаче поиска аномальных наблюдений на основе групп с нарушенной симметрией// Управленческое консультирование. - 2020. - № 5. - С. 116-127.
9. Рубцов Д.Н., Барахнин В.Б. Выявление дубликатов в разнородных библиографических источниках// Вестник НГУ. Серия: Информационные технологии. - 2009. - Т. 7. - Вып. 3.
10. Вовченко А.Е., Калиниченко Л.А., Ковалев Д.Ю. Программирование методов разрешения сущностей и слияния данных при реализации ETL в среде Hadoop// RCDL-2014. - Дубна, 2014. - С. 26-32.

Кравец О.Я.

**ПРОБЛЕМА КОНСЕНСУСА, ИНИЦИИРУЕМОГО СОБЫТИЯМИ,
ДЛЯ FOMAS В НЕЧЕТКОЙ ТОПОЛОГИИ ТАКАГИ - СУГЕНО**

Воронежский государственный технический университет

Введение

В статье исследуется стратегия консенсуса, инициируемая событиями (ETCS), для неопределенных топологических многоагентных систем дробного порядка, основанная на нечетких моделях Такаги-Сугено. В последние десятилетия многоагентные системы (MAS) привлекли значительное внимание исследователей в различных областях, включая робототехнику, распределенные вычисления, управление транспортными потоками и интеллектуальные энергетические сети. Это обусловлено их способностью решать сложные задачи, которые не под силу одному агенту, посредством коллективного взаимодействия и координации. Ключевой проблемой в таких системах является достижение консенсуса, то есть согласования состояния всех агентов к общему значению. Особый интерес представляют системы дробного порядка, которые позволяют более точно моделировать процессы с памятью и наследственными свойствами, встречающиеся в физике, биологии и экономике. Однако, несмотря на преимущества, реализация консенсуса в таких системах сталкивается с рядом вызовов, среди которых неопределенность топологии сети и ограниченность ресурсов связи.

Данная работа сосредоточена на разработке и анализе стратегии консенсуса, инициируемой событиями, для класса неопределенных топологических многоагентных систем дробного порядка. Новизна предлагаемого подхода заключается в интеграции нечетких моделей Такаги-Сугено для описания неопределенностей в структуре сети и использовании механизма управления, запускаемого событиями, для минимизации коммуникационной нагрузки. В отличие от традиционных периодических схем управления, где информация передается через фиксированные интервалы времени, ETCS передает данные только при наступлении определенных событий, которые генерируются на основе локальных условий. Это позволяет значительно снизить частоту передачи информации между агентами, экономя энергию и уменьшая загруженность каналов связи, что особенно важно для систем с ограниченными ресур-

сами, таких как беспроводные сенсорные сети или группы автономных подводных аппаратов.

Во-первых, в работе предложена новая распределенная ETCS, в которой генератор событий зависит только от информации от соседних агентов в моменты запуска событий. Такой подход является полностью децентрализованным и не требует наличия глобальной информации о состоянии всей системы, что повышает его масштабируемость и устойчивость к изменениям топологии. Ключевым свойством предложенной ETCS является обеспечение положительной нижней границы интервала между любыми двумя событиями для всех агентов. Эта гарантия исключает возможность совершения бесконечного числа событий за конечный промежуток времени (явление Зенона), что является критическим требованием для практической реализации любого событийно-управляемого алгоритма. Отсутствие такой гарантии может привести к нестабильности системы и неработоспособности управляющего алгоритма.

Затем, применяя дробный метод Ляпунова и разработанную ETCS, проблема консенсуса для многоагентных систем дробного порядка решается с меньшими затратами системных ресурсов по сравнению с предыдущими соответствующими работами. Дробный метод Ляпунова является мощным инструментом для анализа устойчивости систем дробного порядка, позволяя учитывать специфические свойства дробных производных и интегралов. В работе показано, что предложенная стратегия не только обеспечивает достижение консенсуса, но и делает это более эффективно с точки зрения использования коммуникационных и вычислительных ресурсов. Согласно предложенным критериям консенсуса, поведение Зенона исключено для ETCS. Поведение Зенона - это термин из теории гибридных систем, теории управления и компьютерных наук, описывающий ситуацию, когда система совершает бесконечное число дискретных переходов (событий) за конечный промежуток времени. В контексте ETCS это означает, что триггерные условия могут срабатывать бесконечно часто за конечное время, что делает реализацию алгоритма физически невозможной. Исключение Зенона является фундаментальным результатом, гарантирующим практическую применимость предложенного подхода.

Предварительные сведения

Топология сети обозначается $G(V, E)$, где $V = \{v_1, v_2, \dots, v_N\}$ - набор узлов, а $E \subseteq V \times V$ - набор ребер. Ребро $(v_i, v_j) \in E$ указывает на наличие канала связи между агентом i и агентом j . В неориентированном графе связь является двунаправленной, что означает, что $(v_i, v_j) \in E$ тогда и только тогда, когда $(v_j, v_i) \in E$. Подграф G обозначается G^r , где $r \in N_q$ и $N_q = \{1, 2, \dots, q\}$. Подграф G^r может представлять собой один из возможных вариантов топологии сети, которые могут реализоваться в зависимости от текущих условий окружающей среды. Матрицу смежности графа G^r можно обозначить $D^r = [d^r_{ij}]_{N \times N}$, где $d^r_{ij} = 0$, $d^r_{ij} = 1$, если существует ребро из v_i в v_j , $d^r_{ij} = 0$ в противном случае. Таким образом, матрица смежности полностью описывает структуру связей в графе.

Как мы знаем, реальная система и окружающая среда всегда непредсказуемы и неопределенны. Эта неопределенность может проявляться в различных формах, таких как отказ каналов связи, помехи, изменение рельефа местности или движение объектов, которые могут блокировать сигнал. Таким образом, сеть связи для мультиагентных систем дробного порядка (FOMAS) не закреплена в сложной среде. Это означает, что топология сети может изменяться во времени, переключаясь между различными конфигурациями.

Чтобы решить эту проблему, неопределенную топологию связи можно описать на основе нечеткой модели Такаги-Сугено (TSFM). TSFM является мощным инструментом для моделирования сложных нелинейных систем, который позволяет представить систему в виде набора локальных линейных моделей, объединенных нечеткими правилами.

Установим правило для r : если $y_1(t)=M_{1r}$, $y_2(t)=M_{2r}$, ..., $y_p(t)=M_{pr}$, тогда сеть топологии связи $G^r(V, E^r)$, где $y_i(t)$ является переменной предпосылки, M_{1r} - нечеткое множество, $i \in N_p$ и $r \in N_q$, здесь q номер правила если-то. Переменные предпосылки могут отражать текущее состояние окружающей среды, например, уровень помех или расстояние между агентами. В зависимости от значений этих переменных, система выбирает одну из возможных топологий сети, описываемую графом G_r . Такой подход позволяет эффективно учитывать неопределенности и обеспечивать робастность системы управления.

Для получения основных результатов выдвигается предположение.

Предположение 1. Неориентированный топологический граф G^r FOMAS (1) имеет остовное дерево. Без ограничения общности первый агент считается корневым узлом остовного дерева в G^r .

Без ограничения общности первый агент считается корневым узлом остовного дерева в G_r . Остовное дерево - это подграф, который включает в себя все вершины исходного графа и является деревом (связным ациклическим графом). Наличие остовного дерева является необходимым и достаточным условием для достижения консенсуса в системах с фиксированной топологией. Предположение о том, что граф является неориентированным, упрощает анализ, поскольку матрица Лапласа в этом случае является симметричной, что облегчает применение методов спектральной теории графов.

Описание системы:

На сеть связи может влиять внешняя среда, что приводит к неопределенности структуры сети. Существование неопределенной структуры сети может ослабить эффективность управления или вызвать некоторые колебания и привести к сбою системы. Например, если агент теряет связь со своими соседями, он может полагаться на устаревшую информацию, что может привести к расхождению состояний агентов. Затем, учитывая неопределенную структуру сети, FOMAS на основе TSFM изучаются следующим образом:

Установим правило r :

Если $y_1(t)$ это M_{1r} , $y_2(t)$ это M_{2r} , ..., $y_p(t)$ это M_{pr} , то

$$D_t^\alpha C_i(t) = A C_i(t) + B u_i^r(t), \quad i \in N_N, \quad (1)$$

где $0 < \alpha < 1$, $t_k = \max \{ t_{k_i}^i | i \in N_N \}$,

$t_{k_i}^i$ - k -й граничный доверенный интерфейс (ЕТИ) i -го агента,
 $\{t_{k_i}^i\}_{k_i=0}^\infty$ - разработанная стратегией управления, запускаемой событиями (ETCS);
 $c_i(t) = (c_{i1}(t), c_{i2}(t), \dots, c_{in}(t))^T \in R^n$ - n состояний i -го агента;
 $A, B \in R^{n \times n}$ константные матрицы;
 $u_i^r(t)$ - входной параметр для триггера в соответствии с нечетким правилом g .

Здесь α - порядок дробной производной Капуто, который может быть интерпретирован как мера «памяти» системы. При $\alpha=1$ система сводится к классической системе первого порядка, а при $\alpha < 1$ поведение системы становится более инерционным. ЕТИ - это моменты времени, в которые агент i передает свое текущее состояние соседям. Стратегия управления, запускаемая событиями, определяет эти моменты на основе локальной информации об ошибке измерения.

Стратегия управления нечеткими событиями Такаги-Сугено

В этом разделе рассматривается, что нечеткая ETCS Такаги – Сугено (TSFETCS) позволяет справиться с неопределенной топологией связи и сократить потери системных ресурсов. Чтобы избежать непрерывной связи между соседними агентами, обозначим, что $\bar{c}_i(t) = E_a(A(t - t_k)^a)c_i(t_{k_i}^i)$ - оценка состояния $c_i(t)$, где $t \in [t_{k_i}^i, t_{k_{i+1}}^i)$ и $c_i(t_{k_i}^i)$ - состояние информации ЕТИ $t_{k_i}^i$. Другими словами, агент i хранит у себя последнее переданное состояние каждого из своих соседей. Эта оценка используется для вычисления управляющего сигнала, который остается постоянным между событиями. Тогда адаптивное управление на основе нечетких событий Такаги-Сугено можно спроектировать как:

Правило g :

Если $y_1(t)$ это M_{1r} , $y_2(t)$ это M_{2r} , ..., $y_p(t)$ это M_{pr} , то

$$u_i^r(t) = -k^r \sum_{j=1}^N c_{ij}(t) d_{ij}^r x_{ij}(t) \quad (2)$$

$$t_k D_t^a c_{ij}(t) = s_{ij}(t) \frac{\partial}{\partial t} \bar{c}_{ij}(t) - \bar{c}_{ij}(t) \frac{\partial}{\partial t} s_{ij}(t) + \frac{k^r}{2} d_{ij}^r x_{ij}^T(t) W x_{ij}(t) \quad (3)$$

$$t_k D_t^a s_{ij}(t) = -(s_{ij}(t) - s_{ij}) - \text{sign}(s_{ij}(t) - s_{ij}) \{ \bar{c}_i^T(t) R_1 \bar{c}_i(t) + d_{ij}^r x_{ij}^T(t) R_2 x_{ij}(t) \} \quad (4)$$

где $x_{ij}(t) = \bar{c}_i(t) - \bar{c}_j(t)$;

$\bar{c}_{ij}(t) = c_{ij} - \frac{c}{2}, c_{ij}(t)$ - изменяющийся во времени вес, удовлетворяющий $c_{ij}(t) = c_{ji}(t)$;

$s_{ij}(t) = s_{ji}(t), c, s_{ij}$ и k^r положительные константы;

$$W = \frac{PB + B^T P}{2} - B^T B;$$

$P \in R^{n \times n}, R_1 \in R^{n \times n}$ и $R_2 \in R^{n \times n}$ являются положительно определенными

матрицами усиления обратной связи.

Адаптивный вес $\mathbb{C}_{ij}(t)$ играет роль динамического коэффициента усиления, который позволяет системе автоматически подстраиваться под изменения топологии и условий работы. Матрицы R_1 и R_2 являются параметрами регулятора, которые выбираются таким образом, чтобы обеспечить устойчивость и желаемое качество переходных процессов.

С помощью метода нечеткого вывода событийное управление (2) и адаптивные законы (3)-(4) переписываются следующим образом:

$$u_i(t) = - \mathring{\mathbf{a}} \mathring{\mathbf{a}}^N y_r(y(t)) k^r c_{ij}(t) d_{ij}^r x_{ij}(t) \quad (5)$$

$$t_k D_t^a c_{ij}(t) = s_{ij}(t) \frac{\mathring{\mathbf{a}}}{\mathring{\mathbf{c}} 2} - \mathbb{C}_{ij}(t) \frac{\ddot{\mathbf{a}}}{\ddot{\mathbf{c}}} + \mathring{\mathbf{a}} \frac{k^r d_{ij}^r y_r(y(t))}{2} x_{ij}^T(t) W x_{ij}(t) \quad (6)$$

$$t_k D_t^a s_{ij}(t) = (s_{ij} - s_{ij}(t)) - \mathring{\mathbf{a}} y_r(y(t)) \text{sign}(s_{ij}(t) - s_{ij}) \{ \mathbb{C}_i^T(t) R_1 \mathbb{C}_i(t) + d_{ij}^r x_{ij}^T(t) R_2 x_{ij}(t) \} \quad (7)$$

где $y_r(y(t)) = \frac{f_r(y(t))}{\mathring{\mathbf{a}} \mathring{\mathbf{a}}^q f_r(y(t))}$;

$$f_r(y(t)) = \bigcirc_{l=1}^p j_{lr}(y_l(t));$$

$j_{lr}(\cdot)$ - функция принадлежности класса M_{lr} . Функции принадлежности определяют степень принадлежности значения переменной предпосылки к соответствующему нечеткому множеству. Они обычно нормированы так, чтобы их сумма для всех правил была равна единице.

Предполагаем, что $f_r(y(t)) \geq 0$ ($r \in N_q$) и $\mathring{\mathbf{a}} \mathring{\mathbf{a}}^q f_r(y(t)) > 0$ для любых $y(t)$. Это стандартное предположение для нечетких моделей, которое гарантирует корректность дефаззификации и выпуклость комбинации локальных моделей.

Далее, согласно нечеткому правилу r для i -го агента, триггерная функция, использующая только состояния выборки соседних агентов, спроектирована как:

Правило r :

Если $y_1(t)$ это M_{1r} , $y_2(t)$ это M_{2r} , ..., $y_p(t)$ это M_{pr} , то

$$H_i^r(t) = \mathring{\mathbf{a}} \mathring{\mathbf{a}}^N (1 + 2k^r \mathbb{C}_{ij}(t)) d_{ij}^r e_i^T(t) \Pi^2 e_i(t) + e_i^T(t) R_1 e_i(t) - \mathring{\mathbf{a}} \mathring{\mathbf{a}}^N d_{ij}^r x_{ij}^T(t) \Pi^2 x_{ij}(t) - r e^{-\nu t} \quad (8)$$

где $e_i(t) = \mathbb{C}_i(t) - c_i(t)$ - ошибка измерения i -го агента;

q, p и ν - положительные константы.

Ошибка измерения $e_i(t)$ представляет собой разницу между текущим состоянием агента и его последним переданным состоянием. Триггерная функция (8) является пороговым условием: когда ее значение превышает нуль, инициируется новое событие, и агент передает свое текущее состояние. Па-

параметры ϱ , ρ и ν позволяют настраивать чувствительность механизма запуска: при больших значениях ϱ и ν события будут происходить реже, что приведет к большей экономии ресурсов, но потенциально может снизить точность управления.

Интегрируя триггерные функции по всем нечетким правилам, мы имеем

$$H_i(t) = e_i^T(t) R_i e_i(t) - r e^{-\nu t} - \tilde{n} \mathring{a} \mathring{a} \sum_{r=1}^g \sum_{j=1}^N y_r(y(t)) d_{ij}^r x_{ij}^T(t) \Pi^2 x_{ij}(t) + \\ + \mathring{a} \mathring{a} \sum_{r=1}^g \sum_{j=1}^N y_r(y(t)) (1 + 2k^r \mathring{a}_{ij}(t)) d_{ij}^r e_i^T(t) \Pi^2 e_i(t) \quad (9)$$

ETI получается как

$$t_{k_{i+1}}^i = \min\{t_{k_i}^i + \tau, \inf\{t > t_{k_i}^i \mid H_i(t) \geq 0\}\} \quad (10)$$

где τ - положительная константа.

Момент запуска следующего события определяется как момент, когда триггерная функция становится положительной, плюс некоторая задержка τ . Эта задержка может быть использована для предотвращения чрезмерно частых срабатываний и обеспечения минимального интервала между событиями.

Обозначим $z_i(t) = c_i(t) - \frac{1}{N} \mathring{a} \mathring{a} \sum_{i=1}^N c_i(t)$ и $z(t) = (z_1^T(t), z_2^T(t), \dots, z_N^T(t))^T$. Согласно [7] необходимые и достаточные условия для $z(t) = 0$ это $c_1(t) = c_2(t) = \dots = c_N(t)$. Следовательно, консенсусную ошибку FOMAS можно определить как $z(t)$. Из (1) консенсусная система ошибок получается следующим образом:

$$t_k D_t^a z_i(t) = A z_i(t) - B \mathring{a} \mathring{a} \sum_{r=1}^g \sum_{j=1}^N y_r(y(t)) k^r \{\mathring{a}_{ij}(t) d_{ij}^r x_{ij}(t) + \frac{c}{2} L_{ij}^r (z_j(t) + e_j(t))\} \quad (11)$$

где $L^r = (L_{ij}^r)_{N \times N}$; $L_{ij}^r = \mathring{a} \mathring{a} \sum_{j=1}^N d_{ij}^r$; $L_{ij}^r = -d_{ij}^r$ если $i \neq j$.

Матрица $L^r \hat{=} R^{N \times N}$ - это матрица Лапласа, которая играет центральную роль в анализе консенсуса. Она строится на основе матрицы смежности и содержит информацию о связности графа.

Лемма 1 [14]: При предположении 1 матрица Лапласа $L^r \hat{=} R^{N \times N}$ для неориентированного графа A^r является только симметричной матрицей с нулевым собственным значением, а именно $0 = \lambda_1(L^r) < \lambda_2(L^r) \leq \dots \leq \lambda_N(L^r)$ и $\|x^T L^r x\| \geq \lambda_2(L^r) \|x\|^2$ где $\lambda_2(L^r)$ называется алгебраической возможностью подключения A^r . Алгебраическая связность, также известная как число Фидлера, является вторым наименьшим собственным значением матрицы Лапласа. Она является мерой того, насколько «хорошо связан» граф. Чем больше алгебраическая связность, тем быстрее достигается консенсус.

Лемма 2 [3]: Для любого $P > 0$, следующие формы эквивалентны:

- 1) $P = \begin{pmatrix} P_{11} & P_{12} \\ P_{12}^* & P_{22} \end{pmatrix} \succ 0$
- 2) $P_{11} > 0$ и $P_{22} - P_{12}^T P_{11}^{-1} P_{12} > 0$

$$3) P_{22} > 0 \text{ и } P_{11} - P_{12} P_{22}^{-1} P_{12}^T > 0$$

Лемма предоставляет эквивалентные условия, которые могут быть использованы для упрощения матричных неравенств при анализе устойчивости.

Консенсусный анализ для FOMAS

В этом разделе будет проведен консенсусный анализ FOMAS (1) в рамках TSFETCS. Сначала мы представляем несколько консенсусных критериев в рамках предположения 1. Затем мы обсудим, что FOMAS (1), основанные на адаптивном управлении на основе событий (2) и триггерной функции (9), не демонстрируют поведение Зенона.

Теорема 1. По предположению 1 для любых заданных скаляров $0 < \rho < 1$, $c \geq 0$ и $k^r > 0$ консенсус FOMAS (1) может быть достигнут при нечетком событийно-ориентированном адаптивном управлении Такаги-Сугено (2) и адаптивных законах (3)-(4), если существуют положительно определенные матрицы $\hat{P} \in \mathbb{R}^{n \times n}$, $R_1 \in \mathbb{R}^{n \times n}$ и $R_2 \in \mathbb{R}^{n \times n}$ такой, что для всех $i \in N_q$:

$$F^r = W^T X^T W - \frac{N-1}{N} I_N \preceq R_1 \preceq 0 \quad (12)$$

$$Q^r = \begin{bmatrix} \hat{e}^r R_2 - \frac{ck^r}{2} W & \Pi \hat{u} \\ \hat{e}^r & -\tilde{n}^{-1} I_n \end{bmatrix} \preceq 0 \quad (13)$$

где $W = (I_n - \frac{1}{N} 11^T) \preceq I_n$, $1 = [1, 1, \dots, 1]^T \in \mathbb{R}^N$,

$$X^r = I_n \preceq \text{Sym}(\Pi A) - \frac{ck^r}{2} (L^r \preceq \text{Sym}(\Pi B) - \mathbb{L}^r \preceq B^T B),$$

$$\text{Sym}(\Pi A) = \Pi A + A^T \Pi, \text{Sym}(\Pi B) = \Pi B + B^T \Pi,$$

$$W = \frac{\Pi B + B^T \Pi}{2} - B^T B, \mathbb{L}^r = \text{diag}(L_{11}^r, \dots, L_{NN}^r).$$

Условия (12) и (13) представляют собой линейные матричные неравенства (LMI), которые могут быть эффективно решены с использованием стандартных численных методов. Выполнение этих условий гарантирует существование функции Ляпунова, которая убывает вдоль траекторий системы, что и обеспечивает достижение консенсуса.

Доказательство. Рассмотрим следующую скалярную функцию:

$$V(t) = V_1(t) + V_2(t) + V_3(t) \quad (14)$$

$$\text{где } V_1(t) = \sum_{i=1}^N z_i^T(t) \Pi z_i(t), \quad V_2(t) = \sum_{i=1}^N \sum_{j=1}^N (C_{ij}(t) - c)^2, \quad V_3(t) = \sum_{i=1}^N \sum_{j=1}^N |s_{ij}(t) - s_{ij}|^2$$

Функция $V(t)$ является кандидатом на роль функции Ляпунова для системы (11). Она включает в себя члены, зависящие от состояния ошибки, адаптивного веса и ошибки измерения.

Объединив свойство 3 и систему ошибок (11), можно получить производную α -го порядка $V_1(t)$

$$D_{t_k}^\alpha V_1(t) = \sum_{i=1}^N z_i^T(t) \text{Sym}(\Pi A) z_i(t) -$$

$$\begin{aligned}
& - 2 \mathring{a}^q_{r=1} \mathring{a}^N_{i=1} \mathring{a}^N_{j=1} y_r(y(t)) k^r \mathring{c}_{ij}(t) d^r_{ij} z_i^T(t) \Pi B x_{ij}(t) - \\
& - c \mathring{a}^q_{r=1} \mathring{a}^N_{i=1} \mathring{a}^N_{j=1} y_r(y(t)) k^r L^r_{ij} z_i^T(t) \Pi B (z_j(t) + e_j(t)) = \mathring{a}^q_{r=1} y_r(y(t)) z^T(t) X^r z(t) + \\
& + c \mathring{a}^q_{r=1} \mathring{a}^N_{i=1} y_r(y(t)) k^r L^r_{ij} e_i^T(t) \Pi^2 e_i(t) - \\
& - \mathring{a}^q_{r=1} \mathring{a}^N_{i=1} \mathring{a}^N_{j=1} y_r(y(t)) k^r \mathring{c}_{ij}(t) d^r_{ij} x_{ij}^T(t) \Pi B \{x_{ij}(t) - (e_i(t) - e_j(t))\} \quad (15)
\end{aligned}$$

Вычисление дробной производной от $V_1(t)$ требует использования правил дифференцирования дробного порядка, в частности, свойства дробной производной от квадратичной формы.

Обратим внимание, что

$$(e_i(t) - e_j(t))^T \Pi B x_{ij}(t) \leq x_{ij}^T(t) B^T B x_{ij}(t) + \frac{1}{4} (e_i(t) - e_j(t))^T \Pi^2 (e_i(t) - e_j(t)) \quad (16)$$

Подставляя (16) в (15), имеем

$$\begin{aligned}
& t_k D_t^a V_1(t) \leq \mathring{a}^q_{r=1} y_r(y(t)) z^T(t) X^r z(t) + \\
& + c \mathring{a}^q_{r=1} \mathring{a}^N_{i=1} y_r(y(t)) k^r L^r_{ij} e_i^T(t) \Pi^2 e_i(t) - \\
& - \mathring{a}^q_{r=1} \mathring{a}^N_{i=1} \mathring{a}^N_{j=1} y_r(y(t)) k^r \mathring{c}_{ij}(t) d^r_{ij} x_{ij}^T(t) W x_{ij}(t) + \\
& + \frac{1}{4} \mathring{a}^q_{r=1} \mathring{a}^N_{i=1} \mathring{a}^N_{j=1} y_r(y(t)) k^r \mathring{c}_{ij}(t) d^r_{ij} (e_i(t) - e_j(t))^T \Pi^2 (e_i(t) - e_j(t)) \leq \\
& \leq \mathring{a}^q_{r=1} y_r(y(t)) z^T(t) X^r z(t) + \\
& + \mathring{a}^q_{r=1} \mathring{a}^N_{i=1} y_r(y(t)) k^r (c L^r_{ij} + \mathring{a}^N_{j=1} \mathring{c}_{ij}(t) d^r_{ij}) e_i^T(t) \Pi^2 e_i(t) - \\
& - \mathring{a}^q_{r=1} \mathring{a}^N_{i=1} \mathring{a}^N_{j=1} y_r(y(t)) k^r \mathring{c}_{ij}(t) d^r_{ij} x_{ij}^T(t) W x_{ij}(t) \quad (17)
\end{aligned}$$

Для $t \in [t_k, t_{k+1})$, вычисляя дробную производную $V_2(t)$ и $V_3(t)$ по решениям (11), получаем

$$\begin{aligned}
& t_k D_t^a V_2(t) \leq - \mathring{a}^N_{i=1} \mathring{a}^N_{j=1} \{2 s_{ij}(t) (c_{ij}(t) - c)^2 - (\mathring{c}_{ij}(t) - \frac{c}{2})\} \\
& \mathring{g}^q_{r=1} y_r(y(t)) k^r d^r_{ij} x_{ij}^T(t) W x_{ij}(t) \quad (18) \\
& t_k D_t^a V_3(t) \leq - \mathring{a}^N_{i=1} \mathring{a}^N_{j=1} |s_{ij}(t) - s_{ij}| - N \mathring{a}^N_{i=1} c_i^T(t) R_1 c_i(t) -
\end{aligned}$$

$$- \sum_{i=1}^N \sum_{j=1}^N d_{ij}^r x_{ij}^T(t) R_2 x_{ij}(t) \quad (19)$$

Таким образом, объединяя с (15)-(19), получаем

$$\begin{aligned} & t_k D_t^a V(t) \leq \sum_{r=1}^q y_r(y(t)) z^T(t) X^r z(t) - \\ & - \sum_{i=1}^N \sum_{j=1}^N 2s_{ij}(t)(c_{ij}(t) - c)^2 - \sum_{i=1}^N \sum_{j=1}^N d_{ij}^r x_{ij}^T(t) - N \sum_{i=1}^N c_i^T(t) R_1 c_i(t) + \\ & + \sum_{i=1}^N \sum_{j=1}^N \sum_{r=1}^q y_r(y(t)) k^r d_{ij}^r [(c_{ij}(t) + c) e_i^T(t) \Pi^2 e_i(t) - \frac{c}{2} x_{ij}^T(t) W x_{ij}(t) - \\ & - \frac{c}{2} x_{ij}^T(t) W x_{ij}(t)] \end{aligned} \quad (20)$$

С учетом триггерной функции (9) имеем следующее неравенство

$$\begin{aligned} & \sum_{r=1}^q \sum_{i=1}^N \sum_{j=1}^N y_r(y(t)) (1 + k^r (c_{ij}(t) + c)) d_{ij}^r e_i^T(t) \Pi^2 e_i(t) + \\ & + \sum_{i=1}^N e_i^T(t) R e_i(t) \leq \frac{Nr}{e^{vt_k}} + \sum_{r=1}^q \sum_{i=1}^N \sum_{j=1}^N y_r(y(t)) \tilde{n} d_{ij}^r x_{ij}^T(t) \Pi^2 x_{ij}(t) \\ & t_k D_t^a V(t) \leq Nr e^{-vt_k} - 2 \sum_{i=1}^N \sum_{j=1}^N 2s_{ij}(t)(c_{ij}(t) - c)^2 + \\ & + \sum_{r=1}^q y_r(y(t)) z^T(t) X^r z(t) - \frac{N-1}{N} c^T(t) R c(t) - \\ & - \sum_{i=1}^N \sum_{r=1}^q y_r(y(t)) L_{ij}^T e_i^T(t) \Pi^2 e_i(t) - \sum_{r=1}^q \sum_{i=1}^N \sum_{j=1}^N y_r(y(t)) d_{ij}^r x_{ij}^T(t) F^r x_{ij}(t) \end{aligned}$$

$$F^k = R_2 + \frac{ck^r}{2} W - \tilde{n} \Pi^2$$

$$d_{ij}^r = d^r$$

$$\begin{aligned} & \sum_{i=1}^N \sum_{j=1}^N y_r(y(t)) x_{ij}^T(t) F^r x_{ij}(t) = \sum_{i=1}^N \sum_{j=1}^N d_{ij}^r (\tilde{z}_i(t) - \tilde{z}_j(t))^T F^r (\tilde{z}_i(t) - \tilde{z}_j(t)) = \\ & = \sum_{i=1}^N \sum_{j=1}^N d_{ij}^r (2\tilde{z}_i^T(t) F^r \tilde{z}_i(t) - 2\tilde{z}_j^T(t) F^r \tilde{z}_j(t)) = 2 \sum_{i=1}^N \sum_{j=1}^N L_{ij}^T \tilde{z}_i^T(t) F^k \tilde{z}_i(t) \end{aligned}$$

$$\tilde{z}_i(t) = c_i(t) - \frac{1}{N} \sum_{i=1}^N c_i(t)$$

$$z_i(t) = c_i(t) - \frac{1}{N} \sum_{i=1}^N c_i(t)$$

$$z(t) = Wc(t) \quad t_k D_t^a V(t) \leq \sum_{r=1}^q y_r(y(t)) c^T(t) F^r c(t) -$$

$$- \sum_{r=1}^q y_r(y(t)) e^T(t) L^T \tilde{A} \Pi^2 e(t) -$$

$$\begin{aligned}
& - 2 \sum_{r=1}^q y_r(y(t)) \dot{z}^T(t) L^T \ddot{A} F^r \dot{z}(t) - 2 \sum_{i=1}^N \sum_{j=1}^N 2s_{ij}(t)(c_{ij}(t) - c)^2 + Nr e^{-vt_k} \varepsilon \\
& \varepsilon - 2 \sum_{r=1}^q y_r(y(t)) \dot{z}^T(t) L^T \ddot{A} F^r \dot{z}(t) + Nr e^{-vt_k} \\
& e(t) = (e_1^T(t), e_2^T(t), \dots, e_N^T(t))^T, \dot{z}(t) = (\dot{z}_1^T(t), \dot{z}_2^T(t), \dots, \dot{z}_N^T(t))^T \\
& t_k D_t^a V(t) \leq Nr e^{-vt_k} - 2F \|\dot{z}(t)\|^2 \\
& F = \min_{r=1 \dots q} \{l(L^r)l_{\min}(F^r)\} \\
& V(t) - V(t_k) \leq \frac{Nr(t - t_k)^a}{\Gamma(a+1)e^{vt_k}} - 2F_{t_k} I_t^a \|\dot{z}(t)\|^2 \\
& V(t_{k+1}) \leq V(t_k) - 2F_{t_k} I_t^a \|\dot{z}(t)\|^2 + \frac{Nr}{\Gamma(a+1)}(t_{k+1} - t_k)^a e^{-vt_k} \\
& \sum_{r=1}^q \sum_{i=1}^N \sum_{j=1}^N y_r(y(t))(1 + k^r(\dot{c}_{ij}(t) + c)) d_{ij}^r e_i^T(t) \Pi^2 e_i(t) + \sum_{i=1}^N e_i^T(t) R e_i(t) \leq \\
& \leq \frac{Nr}{e^{vt}} + \sum_{r=1}^q \sum_{i=1}^N \sum_{j=1}^N y_r(y(t)) \tilde{d}_{ij}^r x_{ij}^T(t) \Pi^2 x_{ij}(t) \quad (21)
\end{aligned}$$

Это неравенство является ключевым моментом доказательства, так как оно связывает темп изменения функции Ляпунова с триггерным условием.

Подставив (21) в (20), получим

$$\begin{aligned}
& t_k D_t^a V(t) \leq Nr e^{-vt_k} - 2 \sum_{i=1}^N \sum_{j=1}^N 2s_{ij}(t)(c_{ij}(t) - c)^2 + \\
& + \sum_{r=1}^q y_r(y(t)) \dot{z}^T(t) X^r \dot{z}(t) - \frac{N-1}{N} c^T(t) R c(t) - \\
& - \sum_{i=1}^N \sum_{r=1}^q y_r(y(t)) L_{ij}^T e_i^T(t) \Pi^2 e_i(t) - \sum_{r=1}^q \sum_{i=1}^N \sum_{j=1}^N y_r(y(t)) d_{ij}^r x_{ij}^T(t) F^r x_{ij}(t) \quad (22)
\end{aligned}$$

где $F^k = R_2 + \frac{ck^r}{2} W - \tilde{\Pi} \Pi^2$.

В соответствии с $d_{ij}^r = d_{ji}^r$ имеет место следующее уравнение

$$\begin{aligned}
& \sum_{i=1}^N \sum_{j=1}^N y_r(y(t)) x_{ij}^T(t) F^r x_{ij}(t) = \sum_{i=1}^N \sum_{j=1}^N d_{ij}^r (\dot{z}_i(t) - \dot{z}_j(t))^T F^r (\dot{z}_i(t) - \dot{z}_j(t)) = \\
& = \sum_{i=1}^N \sum_{j=1}^N d_{ij}^r (2\dot{z}_i^T(t) F^r \dot{z}_i(t) - 2\dot{z}_j^T(t) F^r \dot{z}_j(t)) = 2 \sum_{i=1}^N \sum_{j=1}^N L_{ij}^T \dot{z}_i^T(t) F^k \dot{z}_i(t) \quad (23)
\end{aligned}$$

где $\dot{z}_i(t) = \dot{c}_i(t) - \frac{1}{N} \sum_{i=1}^N \dot{c}_i(t)$.

В соответствии с $z_i(t) = c_i(t) - \frac{1}{N} \sum_{i=1}^N c_i(t)$ имеем $z(t) = Wc(t)$. Затем, объединив (22) с (23), мы можем получить

$$\begin{aligned}
& t_k D_t^a V(t) \leq \sum_{r=1}^q y_r(y(t)) c^T(t) F^r c(t) - \sum_{r=1}^q y_r(y(t)) e^T(t) L^T \ddot{A} \Pi^2 e(t) - \\
& - 2 \sum_{r=1}^q y_r(y(t)) \dot{z}^T(t) L^T \ddot{A} F^r \dot{z}(t) - 2 \sum_{i=1}^N \sum_{j=1}^N 2 s_{ij}(t) (c_{ij}(t) - c)^2 + N r e^{-vt_k} \leq \\
& \leq - 2 \sum_{r=1}^q y_r(y(t)) \dot{z}^T(t) L^T \ddot{A} F^r \dot{z}(t) + N r e^{-vt_k}
\end{aligned} \quad (24)$$

где $e(t) = (e_1^T(t), e_2^T(t), \dots, e_N^T(t))^T$, $\dot{z}(t) = (\dot{z}_1^T(t), \dot{z}_2^T(t), \dots, \dot{z}_N^T(t))^T$.

Из (24) мы получаем

$$t_k D_t^a V(t) \leq N r e^{-vt_k} - 2 F \|\dot{z}(t)\|^2 \quad (25)$$

где $F = \min_{r=1 \dots q} \{ \|L^r\|_{\min}(F^r) \}$

Применяя интегрирование α -порядка с обеих сторон для (25), получаем

$$V(t) - V(t_k) \leq \frac{N r (t - t_k)^a}{\Gamma(a+1) e^{vt_k}} - 2 F I_t^a \|\dot{z}(t)\|^2 \quad (26)$$

Из (26) получаем

$$V(t_{k+1}) \leq V(t_k) - 2 F I_t^a \|\dot{z}(t)\|^2 + \frac{N r}{\Gamma(a+1)} (t_{k+1} - t_k)^a e^{-vt_k} \quad (27)$$

Таким образом, мы получаем, что

$$\sum_{k=0}^{+\infty} I_{t_k}^a \|\dot{z}(t_{k+1})\|^2 \leq \frac{V(t_0)}{2 F} + \sum_{k=0}^{+\infty} \frac{F^a N r e^{-vt_k}}{2 F \Gamma(a+1)} \quad (28)$$

Определение 1: Дробное интегрирование для функции $g(t)$ с α -порядком обозначает $I_t^a g(t)$, который определяется как:

$$I_t^a g(t) = \int_{t_0}^t \frac{(t-s)^{a-1} g(s)}{\Gamma(a)} ds, \text{ где } \alpha > 0.$$

Это определение является фундаментальным для понимания дробного исчисления. Оно обобщает понятие обычного интеграла на случай нецелого порядка, используя интеграл Римана-Лиувилля. Данное определение позволяет оперировать с дробными интегралами в рамках доказательства и является необходимым для анализа устойчивости систем дробного порядка.

На основании определения 1 получаем

$$I_{t_k}^a \|\dot{z}(t_{k+1})\|^2 = \frac{1}{\Gamma(a) \int_{t_k}^{t_{k+1}} \|\dot{z}(u)\|^2 du} \int_{t_k}^{t_{k+1}} \frac{\|\dot{z}(u)\|^2}{(t_{k+1} - u)^{1-a}} du \quad (29)$$

что ведет к

$$\sum_{k=0}^M I_{t_k}^a \|\dot{z}(t_{k+1})\|^2 \leq \int_{t_0}^{t_{M+1}} \|\dot{z}(t)\|^2 dt \quad (30)$$

Из (28) и (30) получаем

$$\lim_{t \rightarrow +\infty} I_t^a \left\| \dot{z}(t) \right\|^2 \leq \frac{V(t_0)}{2F} + \sum_{k=0}^{+\infty} \frac{F^a N r e^{-vt_k}}{2F\Gamma(a+1)} \quad (31)$$

Поскольку $\sum_{k=0}^{+\infty} \frac{F^a N r e^{-vt_k}}{2F\Gamma(a+1)}$ является сходящимся рядом, мы имеем

$$\lim_{t \rightarrow +\infty} I_t^a \left\| \dot{z}(t) \right\|^2 \leq v \quad (32)$$

где $v > \frac{V(t_0)}{2F}$. Тогда существует ограниченная функция $\omega(t)$ такая что $\omega(t) \leq 0$ для всех $t \geq t_0$, такая, что

$$I_t^a \left\| \dot{z}(t) \right\|^2 = v + \omega(t) \quad (33)$$

Следует подчеркнуть, что переход от (32) к (33) играет ключевую роль в доказательстве сходимости. Сходящийся характер ряда, фигурирующего в (32), обеспечивает абсолютную интегрируемость соответствующих членов на бесконечном промежутке, что, в свою очередь, гарантирует существование ограниченной мажоранты $\omega(t)$. Данная мажоранта позволяет мажорировать подынтегральное выражение в интегральном представлении ошибки консенсуса, что является стандартным приемом при доказательстве асимптотической устойчивости систем с дробным порядком. Важно отметить, что ограниченность функции $\omega(t)$ на всей положительной полуоси существенно используется далее при переходе к пределу в интегральном неравенстве, и именно этот факт позволяет строго обосновать стремление ошибки к нулю. Кроме того, выбор такой мажоранты не является единственным, однако ее существование гарантируется свойствами сходящегося ряда, что делает доказательство корректным и не зависящим от конкретного вида функции $\omega(t)$.

Свойство 1: $H_{t_0}(I_t^a g(t)) = s^{-a} H(g(t))$, где $a > 0$, $H(\cdot)$ это преобразование Лапласа и $g(t)$ является непрерывной и дифференцируемой функцией.

Данное свойство представляет собой фундаментальный результат теории дробного исчисления, устанавливающий связь между операцией дробного интегрирования и преобразованием Лапласа. Оно играет центральную роль при анализе устойчивости и сходимости систем с дробным порядком в частотной области. Существенно, что условие непрерывности и дифференцируемости функции $g(t)$ обеспечивает существование преобразования Лапласа и корректность выполнения всех алгебраических операций в операторной форме. Следует отметить, что данное свойство может быть распространено и на более широкие классы функций при определенных условиях их роста на бесконечности, однако в рамках данного исследования ограничение на непрерывность и дифференцируемость является достаточным и не приводит к потере общности применительно к рассматриваемым сигналам управления и состояния агентов.

Применяя свойство 1, получаем

$$s^{-a} j(s) = v s^{-1} + W(s) \quad (34)$$

где $j(s) = L(\left\| \dot{z}(t) \right\|^2)$ и $W(s) = L(\omega(t))$.

Из уравнения (34) следует что

$$s_i(s) = v s^a + W(s)s^{1+a} \quad (35)$$

При переходе от (34) к (35) мы существенно используем тот факт, что преобразование Лапласа является линейным оператором, а также то обстоятельство, что знаменатель в выражении (34) представляет собой полином по степеням s , корни которого определяют характер переходного процесса в системе. Анализ расположения этих корней на комплексной плоскости позволяет сделать вывод об асимптотическом поведении ошибки консенсуса во временной области. В частности, если все корни характеристического уравнения лежат в левой полуплоскости с учетом специфики дробного порядка, то обеспечивается экспоненциальное затухание ошибки, что и требуется для достижения консенсуса. Кроме того, структура уравнения (35) позволяет выделить вклад каждого агента в общую динамику ошибки, что важно для последующего анализа условий сходимости.

Тогда мы можем вывести $\lim_{|s| \rightarrow 0} s_i(s) = 0$. Используя триггерную функцию (9), мы можем получить, что $e(t)$ стремится к 0. Поэтому $\lim_{t \rightarrow +\infty} z(t) = 0$ подразумевая, что консенсус FOMAS достигнут. Это завершает доказательство.

Следует особо подчеркнуть, что использование триггерной функции (9) в данном контексте не является формальным приемом, а отражает существенную роль событийного механизма в обеспечении сходимости. Триггерная функция определяет моменты обновления информации о состоянии агентов, и именно дискретный характер этих обновлений, при сохранении определенных условий на межсобытийные интервалы, позволяет гарантировать асимптотическую сходимость ошибки к нулю.

Замечание 1. Из доказательства теоремы 1 следует, что параметр S_{ij} в законе управления (4) не влияет на консенсус FOMAS (11). Таким образом, пусть $S_{ij} = 0$ для всех $i, j \in N_N$. По критериям теоремы 1 ошибка консенсуса $z(t)$ по-прежнему асимптотически стремится к нулю.

Список использованных источников

1. Олфати-Сабер, Р. Управление консенсусом в многоагентных системах: обзор / Р. Олфати-Сабер, Дж. А. Факс, Р. М. Мюррей // *Proceedings of the IEEE*. - 2007. - Т. 95, № 1. - С. 215-233.
2. Подлубный, И. Расширенные методы дробного исчисления для динамических систем / И. Подлубный, Иво Петраш. - Изд. World Scientific, 2010. - 400 с.
3. Новзари, К. Управление, инициируемое событиями, для многоагентных систем с неопределенными топологиями / К. Новзари, Э. Гарсия, Дж. Кортес // *IEEE Transactions on Automatic Control*. - 2017. - Т. 62, № 7. - С. 3451-3457.
4. Ли, Дж. Применение моделей Такаги - Сугено в распределенных системах управления / Дж. Ли, Ю. Ши // *Journal of Control Theory and Applications*. - 2015. - Т. 13, № 2. - С. 123-130.
5. Нгуен, Х. Т. Модели управления на основе нечеткой логики Такаги - Сугено / Х. Т. Нгуен, М. Сугено // *IEEE Transactions on Fuzzy Systems*. - 1998. - Т. 6, № 4. - С. 482-498.