

На правах рукописи



АТЛАСОВ Денис Игоревич

**УПРАВЛЕНИЕ ПРОЦЕССАМИ ОБРАБОТКИ ГЕТЕРОГЕННЫХ
ДАНЫХ В РАМКАХ ИНФОРМАЦИОННЫХ СИСТЕМ С
МНОГОМЕРНЫМИ АТРИБУТАМИ НА ОСНОВЕ МЯГКОЙ
МАКСИМИННОЙ ОЦЕНКИ**

Специальность: 2.3.5. Математическое и программное обеспечение
вычислительных систем, комплексов и
компьютерных сетей

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата технических наук

Воронеж – 2026

Работа выполнена в ФГБОУ ВО «Воронежский государственный технический университет».

Научный руководитель: **Кравец Олег Яковлевич**, доктор технических наук, профессор

Официальные оппоненты: **Ломакина Любовь Сергеевна**, доктор технических наук, профессор, ФГБОУ ВО «Нижегородский государственный технический университет им. Р.Е. Алексеева», профессор кафедры «Вычислительные системы и технологии»

Смирнова Мария Сергеевна, доктор технических наук, доцент, ФГБОУ ВО "Балтийский государственный технический университет «ВОЕНМЕХ» им. Д.Ф.Устинова", профессор кафедры Н2 "Программная инженерия и интеллектуальные системы"

Ведущая организация: ФГБОУ ВО «Самарский государственный технический университет» (г. Самара)

Защита состоится «17» апреля 2026 года в 14 часов в конференц-зале на заседании диссертационного совета 24.2.286.04, созданного на базе ФГБОУ ВО «Воронежский государственный технический университет», по адресу: г. Воронеж, Московский просп., д. 14, ауд. 216.

С диссертацией можно ознакомиться в научно-технической библиотеке ФГБОУ ВО «Воронежский государственный технический университет» и на сайте <https://cchgeu.ru>.

Автореферат разослан «06» марта 2026 г.

Ученый секретарь
диссертационного совета



Гусев Константин Юрьевич

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. Сложные и разнообразные программные системы требуют постоянного анализа циркулирующих в них данных, так как именно данные определяют полезность и значимость результатов работы программных систем. Рекомендательные системы на основе гетерогенных информационных сетей обеспечивают единый подход к объединению различной вспомогательной информации, которую можно комбинировать с основными алгоритмами рекомендаций для эффективного повышения производительности. Актуальна и проблема извлечения общего сигнала из разнородных данных. Поскольку гетерогенность преобладает в крупномасштабных системах, цель - эффективный в вычислительном отношении оценщик с хорошими статистическими свойствами при различной степени неоднородности данных. Большой вклад в разработку методов и средств управления гетерогенными данными больших вычислительных систем и сетей внесли Бостром Н., Лекун Я., Маркус Г., Bishop С., Goodfellow I.

Одной из актуальных предметных областей задач управления данными гетерогенных систем является управление знаниями. Размер зерна знаний в приближенном пространстве напрямую влияет на неопределенность приближительного множества. С точки зрения интуитивного понимания, чем больше объем знаний, тем меньше информации, тем больше будет неопределенность; чем меньше объем знаний, тем больше информации, тем меньше будет неопределенность. Это центр исследований в области искусственного интеллекта. Важным является измерение неопределенности для гетерогенной информационной системы, которое отражает способность этой системы к классификации и повышению точности классификации данных.

Таким образом, актуальность темы диссертационного исследования продиктована необходимостью разработки моделей и алгоритмов управления процессами обработки гетерогенных данных в рамках информационных систем с многомерными атрибутами на основе мягкой максиминной оценки и активного обучения.

Тематика диссертационной работы соответствует научному направлению ФГБОУ ВО «Воронежский государственный технический университет» «Вычислительные комплексы и проблемно-ориентированные системы управления».

Целью работы является разработка моделей и алгоритмов управления процессами обработки гетерогенных данных в рамках информационных систем с многомерными атрибутами на основе мягкой максиминной оценки и активного обучения.

Задачи исследования. Для достижения поставленной цели необходимо решить следующие задачи:

1. Провести анализ проблем управления гетерогенными данными информационных систем с многомерными атрибутами на основе мягкой максиминной оценки и активного обучения.

2. Разработать мягкую максиминную оценку для гетерогенных данных, содержащих уникальные вариационные компоненты, обеспечивающую со-

хранение статистических свойств и лучшую вычислительную эффективность

3. Предложить архитектуру системы управления гетерогенной информационной системой, обеспечивающую эффективную редукцию многомерных анализируемых атрибутов.

4. Создать алгоритм обнаружения наилучшей модели машинного обучения для анализа гетерогенных данных, обеспечивающий сокращение объемов данных и повышение точности анализа совокупности данных.

5. Разработать алгоритм идентификации нулевых значений в гетерогенных базах данных, обеспечивающий более эффективную и точную оценку нулевых значений.

Объект исследования: процессы управления гетерогенными данными информационных систем с многомерными атрибутами.

Предмет исследования: структура математического и программного обеспечения процессов управления гетерогенными данными информационных систем с многомерными атрибутами.

Методы исследования. При решении поставленных в диссертации задач использовались методы теории вероятностей, теории принятия решений, а также методы объектно-ориентированного программирования.

Тематика работы соответствует следующим пунктам паспорта специальности 2.3.5 «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей»: п. 4. «Интеллектуальные системы машинного обучения, управления базами данных и знаний, инструментальные средства разработки цифровых продуктов»; п. 9. Модели, методы, алгоритмы, облачные технологии и программная инфраструктура организации глобально распределенной обработки данных.

Научная новизна работы. В диссертации получены следующие результаты, характеризующиеся научной новизной:

- мягкая максиминная оценка для гетерогенных данных, содержащих уникальные вариационные компоненты, отличающаяся извлечением надежных данных из разнородных групп, и обеспечивающая сохранение статистических свойств и лучшую вычислительную эффективность;

– архитектура программной системы управления гетерогенной информационной системой, отличающаяся использованием отношения эквивалентности на множестве объектов для измерения неопределенности системы, и обеспечивающая редукцию многомерных анализируемых атрибутов на основе грануляции информации и информационной энтропии;

– алгоритм выбора наилучшей модели машинного обучения для анализа гетерогенных данных, отличающийся применением активного обучения и ансамблевых методов для классификаторов при решении проблем анализа разнородных данных, обеспечивающий сокращение объемов данных и повышение точности анализа совокупности данных;

- алгоритм идентификации нулевых значений в гетерогенных базах данных, отличающийся предварительной классификацией исходных данных на основе взвешенных значений и обеспечивающий более эффективную и точную оценку нулевых значений в среднем на 7.6%.

Теоретическая и практическая значимость исследования заключается в разработке моделей и алгоритмов управления гетерогенными данными информационных систем с многомерными атрибутами на основе мягкой максиминной оценки и активного обучения.

Теоретические результаты работы могут быть использованы в проектных и научно-исследовательских организациях, занимающихся проектированием программных систем с гетерогенными базами данных с многомерными атрибутами.

Положения, выносимые на защиту

1. Мягкая максиминная оценка для гетерогенных данных, содержащих уникальные вариационные компоненты, сохраняет статистические свойства наборов данных и обеспечивает лучшую вычислительную эффективность по сравнению с максиминной оценкой.

2. Архитектура программной системы управления гетерогенной информационной системой реализует эффективную редукцию многомерных анализируемых атрибутов на основе грануляции информации и информационной энтропии.

3. Алгоритм выбора наилучшей модели машинного обучения для анализа гетерогенных данных обеспечивает сокращение объемов анализируемых данных и повышение точности анализа совокупности данных.

4. Алгоритм идентификации нулевых значений в гетерогенных базах данных обеспечивает более эффективную и точную оценку нулевых значений в среднем на 7.6%.

Результаты внедрения. Основные результаты внедрены в ООО М-Сервис (г. Воронеж) при проектировании систем управления гетерогенными программными системами, в учебный процесс Воронежского государственного технического университета в рамках дисциплин: «Вычислительные машины, системы и сети», «Информационные сети и телекоммуникационные технологии», а также в рамках курсового и дипломного проектирования.

Апробация работы. Основные положения диссертационной работы докладывались и обсуждались на следующих конференциях: XXVIII-th и XXX-th International Open Science Conference «Modern informatization problems in simulation and social technologies» (Yelm, WA, USA, 2023, 2025); Международной молодежной научной школе «Оптимизация и моделирование в автоматизированных системах» (Воронеж, 2023); XXIX-th International Open Science Conference «Modern informatization problems in the technological and telecommunication systems analysis and synthesis» (Yelm, WA, USA, 2024); Международной научно-практической конференции, «Интеллектуальные информационные системы» (Воронеж, 2024); VI Всероссийской научно-практической конференции «Информационные технологии в экономике и управлении» (Махачкала, 2024), а также на научных семинарах кафедры автоматизированных и вычислительных систем ВГТУ (2023-2026 гг.).

Достоверность результатов обусловлена корректным использованием теоретических методов исследования и подтверждена результатами сравнительного анализа данных вычислительных и натуральных экспериментов.

Публикации. По результатам диссертационного исследования опубликовано 19 научных работ, в том числе 7 – в изданиях, рекомендованных ВАК РФ (из них 1 – в издании, индексируемых в WoS и одно свидетельство о регистрации программы для ЭВМ). В работах, опубликованных в соавторстве и приведенных в конце автореферата, лично автором получены следующие результаты: [2, 3, 8, 10, 11] - мягкая максиминная оценка для гетерогенных данных, содержащих уникальные вариационные компоненты, отличающаяся извлечением надежных данных из разнородных групп, и обеспечивающая сохранение статистических свойств и лучшую вычислительную эффективность; [1, 6, 7] - архитектура программной системы управления гетерогенной информационной системой, отличающаяся использованием отношения эквивалентности на множестве объектов для измерения неопределенности системы, и обеспечивающая редукцию множества анализируемых атрибутов; [5, 12, 13, 15, 17] - алгоритм выбора наилучшей модели машинного обучения для анализа гетерогенных данных, отличающийся применением активного обучения и ансамблевых методов для классификаторов при решении проблем анализа разнородных данных, обеспечивающий сокращение объемов данных и повышение точность анализа совокупности данных; [4, 9, 16, 18] - алгоритм идентификации нулевых значений в гетерогенных базах данных, отличающийся предварительной классификацией исходных данных на основе взвешенных значений и обеспечивающий более эффективную и точную оценку нулевых значений.

Структура и объем работы. Диссертационная работа состоит из введения, четырех глав, заключения, списка литературы из 208 наименований. Работа изложена на 179 страницах.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность исследования, сформулированы его цель и задачи, научная новизна и практическая значимость полученных результатов, приведены сведения об апробации и внедрении работы.

В первой главе исследуются проблемы управления гетерогенными данными информационных систем с многомерными атрибутами на основе мягкой максиминной оценки и активного обучения. Отмечено, что повысить эффективность такого управления можно путем применения мягкой максиминной оценки для гетерогенных данных, содержащих уникальные вариационные компоненты, архитектуры системы управления гетерогенной информационной системой с редукцией многомерных анализируемых атрибутов, выбора наилучшей модели машинного обучения для анализа гетерогенных данных, создания алгоритмов идентификации нулевых значений в гетерогенных базах данных. Результат анализа потребовал формализации данных задач, а также алгоритмизации их решения с учетом особенностей, отраженных на рис. 1. Сформулирована цель и задачи исследования.

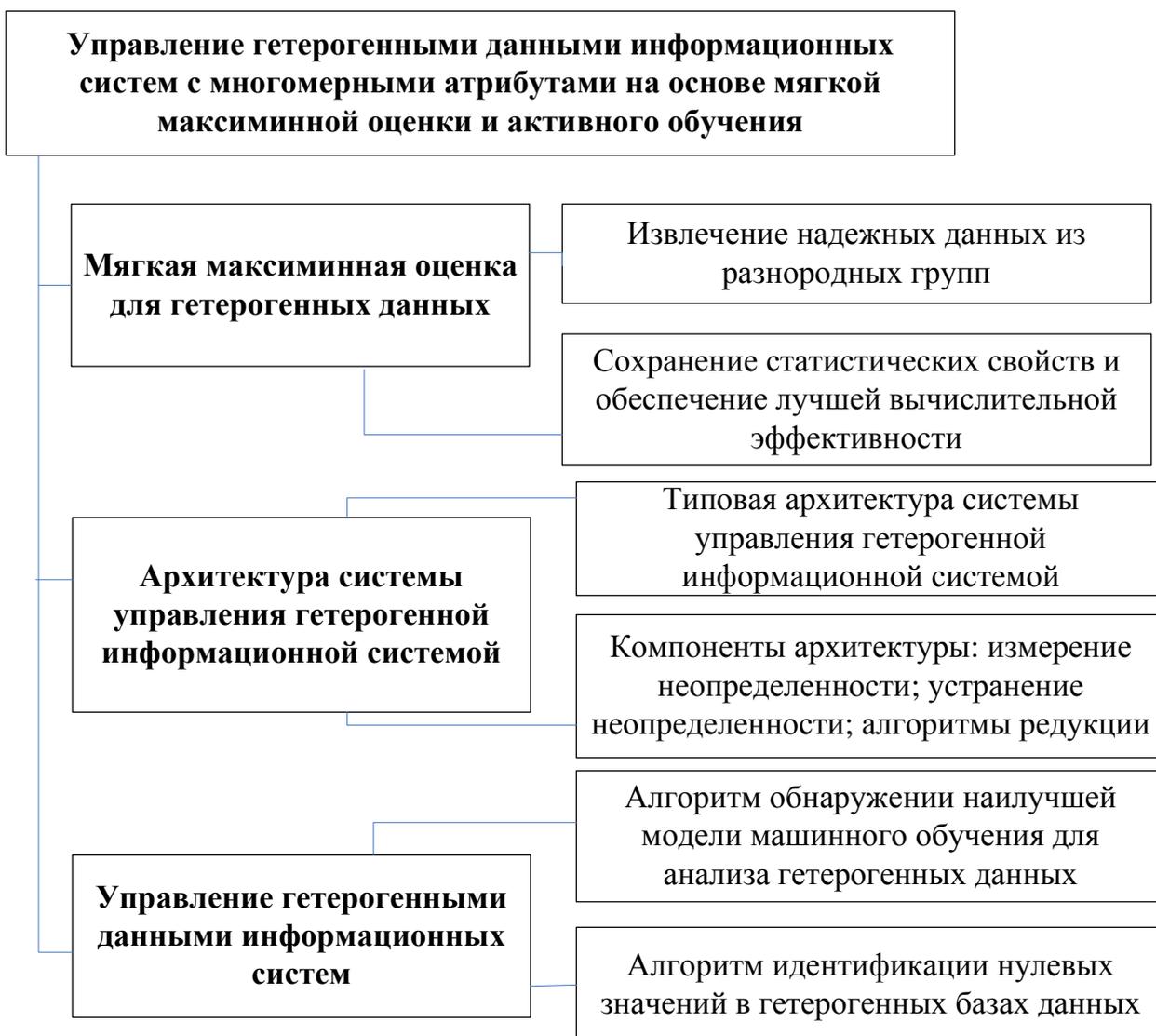


Рис. 1. Дизайн исследования

Вторая глава посвящена развитию инструментов интерполяции мягкой максиминной оценки для гетерогенных данных.

Извлечение общего надежного сигнала из данных, разделенных на разнородные группы, является сложной задачей, когда каждая группа - в дополнение к сигналу - содержит большие уникальные вариационные компоненты. Ранее максиминная оценка была предложена в качестве надежного метода при наличии неоднородного шума. Предлагается мягкая максиминная оценка максимального значения в качестве привлекательной с вычислительной точки зрения альтернативы, направленной на достижение баланса между объединенной оценкой и (жесткой) оценкой максимального значения. Метод мягкой максиминной оценки предоставляет диапазон оценок, управляемых параметром $\xi > 0$, который интерполирует объединенную оценку наименьших квадратов и максиминную оценку. Устанавливая соответствующие теоретические свойства, утверждается, что метод мягкой максиминной оценки является статистически обоснованным и привлекательным с точки зрения вычислений. Демонстрируется на реальных и смоделированных данных, что мягкая максиминная оценка может предложить улучшения как по сравнению с объ-

единенными методами наименьших квадратов (МНК), так и по сравнению с жестким максимумом с точки зрения производительности прогнозирования и вычислительной сложности.

Пусть Y_1, \dots, Y_n - одномерными переменными отклика

$$Y_i = X_i^T B_i + \varepsilon_i, \quad i = 1, \dots, n. \quad (2.1)$$

Здесь $\varepsilon_1, \dots, \varepsilon_n$ и векторы признаков X_1, \dots, X_n являются p -мерными случайными величинами, а B_1, \dots, B_n - одномерными шумовыми переменными. Векторы признаков наблюдаются и предполагаются независимыми одинаково распределёнными случайными величинами, а шумовые переменные равны и предполагаются независимыми одинаково распределёнными случайными величинами. ненаблюдаемые переменные B_1, \dots, B_n распределены идентично распределению F_B , но не обязательно независимы.

Неоднородность в модели, приведенной в (2.1), обусловлена изменением смещения, регулируемым F_B . Поскольку B_i может быть зависимым, модель (2.1) может фиксировать неоднородность, вызванную структурой группы.

Цель настройки с группами G с фиксированными внутригрупповыми B_i состоит в том, чтобы изучить один $\beta \in \mathbb{R}^p$, который можно разумно рассматривать как общий сигнал B_i . Объединение данных по группам и вычисление обычной оценки МНК может быть ненадежным, в зависимости от F_B , и введена максиминная оценка в качестве надежной альтернативы МНК для разнородных данных из модели (2.1).

Хотя максиминная оценка надежна, она также может быть консервативной, и предлагается мягкая максиминная оценка, чтобы обеспечить хороший баланс между максиминной оценкой и объединенной оценкой МНК. Баланс контролируется параметром настройки $\xi > 0$, при этом $\xi \rightarrow \infty$ соответствует максиминной оценке.

Пусть в (2.1) $G \in \mathbb{N}$ групп. Определим I_1, \dots, I_G множества $\{1, \dots, n\}$ так, что $|I_g| = n_g$, а $n = \sum_g n_g$, где $\{g, i \mid I_g: B_i = B_g\}$ в g . Тогда

$$\text{supp}(F_B) = \{b_1, \dots, b_G\} \subset \mathbb{R}^p, \quad b_g := B_g(\omega).$$

Пусть $Y_g = (Y_{g,1}, \dots, Y_{g,n_g})^T$ - вектор отклика $n_g \times 1$, $X_g = (X_{g,1}, \dots, X_{g,n_g})^T$ - след $n_g \times p$, $e_g = (e_{g,1}, \dots, e_{g,n_g})^T$ - вектор ошибки $n_g \times 1$. Тогда

$$Y_g = X_g b_g + e_g, \quad g \in \{1, \dots, G\}. \quad (2.2)$$

Общий сигнал в этой структуре представлен как $b \in \mathbb{R}^p$ так, что X_g является хорошей и надежной аппроксимацией $X_g b_g$ во всех группах G .

Чтобы оценить качество аппроксимации, мы используем критерий оптимальности. Объясненная дисперсия в группе g при использовании некоторого $b \in \mathbb{R}^p$ в (2.2) определяется как

$$V_{bg}(b) = 2b^T \Sigma b_g - b^T \Sigma b. \quad (2.3)$$

Итак, оптимальным приближением для всех групп является максимина $b^* \in \mathbb{R}^p$, который максимизирует минимум объясненных различий между группами, т.е.

$$b^* = \arg \max_b \min_g V_{b_g}(b).$$

Пусть $\hat{\mathbf{a}}_g = \mathbf{X}_g^T \mathbf{X}_g / n_g$ - эмпирическая матрица Грама в группе g . Заменив Σ на Σ_g в (2.3), получим эмпирически объясненную дисперсию в группе g :

$$V_g(b) := (2b^T \Sigma b_g - b^T \Sigma b) / n_g.$$

$$\hat{V}_g(b) := \frac{1}{n_g} (2b^T \mathbf{X}_g^T \mathbf{Y}_g - b^T \mathbf{X}_g^T \mathbf{X}_g(b)) \quad (2.4)$$

В крупномасштабных системах обработки данных, где обычно встречается неоднородность данных, вычислительный аспект оценки имеет решающее значение.

Устраним это вычислительное препятствие, заменив функцию максимума гладкой функцией. Для $G \in \mathbb{N}$ и $z \neq 0$ рассмотрим масштабированную экспоненциальную функцию лсе логарифмической суммы

$$lse_z(x) = \frac{\log(\hat{\mathbf{a}}_j e^{zx_j})}{z}, \quad x \in \mathbb{R}^G \quad (2.5)$$

Предположение 2.1. Пусть $G \in \mathbb{N}$ и $x \in \mathbb{R}^G$. Тогда

При $z > 0$

$$\max(x_1, \dots, x_G) \leq lse_z(x) \leq \frac{\log(G)}{z} + \max(x_1, \dots, x_G) \quad (2.6)$$

и в частности $lse(x) \searrow \max_g \{x_g\}$ при $z \rightarrow \infty$.

При $z \rightarrow 0$

$$lse_z(x) = \frac{1}{G} \sum_{j=1}^G x_j + \frac{\log(G)}{z} + o(1)$$

Определим мягкую функцию максимальных потерь как

$$s_z(b) = lse_z(-\hat{V}(b)), \quad b \in \mathbb{R}^p, \quad z > 0$$

где $\hat{V}(b) = (\hat{V}_1(b), \dots, \hat{V}_G(b))^T$. Для $k > 0$ и $z > 0$, оценка мягкого максимума теперь может быть определена как

$$\hat{b}_{s_{mm}}^k = \arg \min_b lse_z(-\hat{V}(b)) \quad \text{такое, что } \|b\|_1 \leq k \quad (2.7)$$

Следующий результат дает оценку максимальной отрицательной объясненной дисперсии оценки мягкого максимина, используя оценку теоретического эффекта максимина b^* .

Утверждение 2.1. Пусть $D = \max_g \|\hat{\mathbf{S}}_g - \mathbf{S}\|_{\Psi}$ и $d = \max_g \|\mathbf{X}_g^T \mathbf{e}_g\|_{\Psi}$. Тогда

$$\max_g \left\{ -V_{b_g}(\hat{b}_{s_{mm}}^k) \right\} \leq \max_g \left\{ -V_{b_g}(b^*) \right\} + 6Dk^2 + 4kd + \frac{\log(G)}{z}$$

при $z > 0, k > 0, k \geq \max_g \|b_g\|_1$

где b^* - максимальный эффект. В частности

$$\|\hat{b}_{\text{smm}}^k - b^*\|_S \leq 6Dk^2 + 4kd + \frac{\log(G)}{z}$$

Утверждение 2.1 устанавливает связь между производительностью мягкого максимина и эффектом максимина и показывает, что для $z \uparrow \infty$ мы действительно получаем максиминную оценку производительности.

Однако это также подчеркивает, что при $z \downarrow 0$ производительность мягкой оценки максимина может сколь угодно сильно отклоняться от производительности оценки максимина. Действительно, согласно Предположению 2.1, для малых $z > 0$,

$$s_z(b) \gg -\frac{1}{G} \hat{a}_{i=1}^G \hat{V}_j(b) + \frac{\log(G)}{z} \gg \frac{1}{n} \hat{a}_{j=1}^G \hat{a}_{i=1}^{n_j} \frac{n}{Gn_j} \left((X_j b)_i - Y_{j,i} \right)^2$$

и (2.7) фактически становится задачей взвешенных наименьших квадратов со штрафом (PWLS) по всем n наблюдениям. Таким образом, решение (2.7) для малого $z=0$ приблизительно дает объединенную оценку WLS с весами, усиливающими наблюдения из групп меньшего размера, чем в среднем. При одинаковом количестве наблюдений в каждой группе программа оценки мягкого максимума, в свою очередь, интерполирует объединенную оценку PLS и максимальную оценку.

Общая мягкая функция максимальных потерь $l_\zeta: \mathbb{R}^p \rightarrow \mathbb{R}$ теперь задается формулой

$$l_x(b) = \text{lse}_x \mathbf{g}(b) = \frac{\log \hat{a}_{j=1}^G e^{x h_j(b)} \hat{a}_{i=1}^G}{x}, x > 0$$

и цель состоит в том, чтобы решить общую задачу мягкого максимина, сформулированную следующим образом

$$\min_{b \in \mathbb{R}^p} (l_x(b) + l J(b)), l \geq 0 \tag{2.8}$$

Здесь J - выпуклая штрафная функция, а l - штрафной параметр.

Решение (2.8) в крупномасштабных условиях требует эффективного алгоритма оптимизации для недифференцируемых задач. В отличие от задачи жесткого максимина, (2.8) является, в дополнение к выпуклой и недифференцируемой, (частично) дифференцируемой и разделимой задачей. Это означает, что ряд эффективных алгоритмов решит задачу (2.8), например, алгоритмы разделения операторов первого порядка, такие как ADMM, или алгоритм второго порядка, такой как координатный спуск.

Таким образом, разработана мягкая максиминная оценка для гетерогенных данных, содержащих уникальные вариационные компоненты, отличающаяся извлечением надежных данных из разнородных групп, и обеспечивающая сохранение статистических свойств и лучшую вычислительную эффективность.

Третья глава посвящена измерению неопределенности гетерогенных данных и редукции атрибутов в гетерогенных информационных системах.

Исследуется измерение неопределенности для разнородных данных и приводится его применение для редукции атрибутов. Сначала предлагается концепция гетерогенной информационной системы (HIS).

Затем строится отношение эквивалентности на множестве объектов. Затем исследуется измерение неопределенности для HIS, проводится численный эксперимент, в котором проведен дисперсионный анализ, корреляционный анализ, а также тест Фридмана и тест Бонферрони–Данна в статистике. В качестве применения предложенных мер изучается уменьшение атрибутов в HIS, и предлагаются соответствующие алгоритмы и их анализ.

Неопределенность присутствует повсюду, и информационная система не является исключением. Гетерогенная информационная система (HIS) или информационная система с гетерогенными данными означает информационную систему, наборы данных которой содержат данные трех типов (т.е. масштабированные типы, упорядоченные типы и обычные типы). Кроме того, гетерогенная информационная система обладает неопределенностью. Измерение неопределенности для гетерогенной информационной системы отражает способность этой системы к классификации, которая оказывает существенное влияние на точность классификации данных.

Приведем ряд определений и утверждений о свойствах HIS.

Определение 3.1. Пусть U - конечный набор объектов. Предположим, что A выражает конечный набор атрибутов. Тогда упорядоченная пара (U, A) называется информационной системой, если $a \in A$ может определять функцию $a: U \rightarrow V_a$, где $V_a = \{a(u) : u \in U\}$.

Определение 3.2. (U, A) называется гетерогенной информационной системой (HIS) или информационной системой с гетерогенными данными, если ее набор данных содержит три типа данных (т.е. масштабируемые типы, упорядоченные типы и обычные типы).

Далее считаем, что (U, A) - HIS, $B \subseteq A$.

Определение 3.3. Детализация информации в подсистеме (U, B) определяется следующим образом

$$G(B) = - \frac{1}{n^2} \overset{\circ}{\mathbf{a}} \sum_{i=1}^m |X_i|^2, \text{ где } U \setminus R_B^* = \{X_1, X_2, \dots, X_m\}.$$

$$\text{Утверждение 3.1. } G(B) = \frac{1}{n^2} \overset{\circ}{\mathbf{a}} \left| [X_i]_{R_b^*} \right|$$

Детализация информации в HIS полностью основана на информационной структуре этого нечеткого отношения.

$$\text{Утверждение 3.2. } \frac{1}{n} \leq G(B) \leq 1.$$

Более того, если $R_B^* = \Delta$, то $G(B)$ достигает минимального значения $\frac{1}{n}$; если $R_B^* = \omega$, то $G(B)$ достигает максимального значения 1.

Утверждение 3.2 дает диапазон значений грануляции информации для неопределенности HIS.

Утверждение 3.3. $G(Q) \leq G(P)$.

Детализация информации соответствует характеристике грануляционных вычислений и уточняет информацию с разных уровней. Этот показатель монотонно уменьшается по мере уточнения эквивалентного класса. Недостаток заключается в том, что этот показатель не позволяет отличить два обращения с одинаковой детализацией информации, но разными информационными структурами.

Определение 3.4. Информационная энтропия подсистемы (U, V) определяется следующим образом:

$$H(R) = - \sum_{i=1}^m \frac{|X_i|}{n} \log_2 \frac{|X_i|}{n}, \text{ где } U \setminus R_B^* = \{X_1, X_2, \dots, X_m\}.$$

Утверждение 3.5. $H(B) = - \sum_{i=1}^m \frac{1}{n} \log_2 \frac{|[X_i]_{R_B^*}|}{n}$

Утверждение 3.6. $0 \leq H(B) \leq \log_2 n$.

Более того, если $R_B^* = \Delta$, то $H(B)$ достигает максимального значения $\log_2 n$; если $R_B^* = \omega$, то $H(B)$ достигает минимального значения 0.

Утверждение 3.7. Если $P \subseteq Q \subseteq A$, то $H(P) \leq H(Q)$.

Определение 3.5. Грубая энтропия подсистемы (U, V) определяется следующим образом:

$$E_r(B) = - \sum_{i=1}^m \frac{|X_i|}{n} \log_2 \frac{1}{|X_i|}, \text{ где } U \setminus R_B^* = \{X_1, X_2, \dots, X_m\}$$

Утверждение 3.8. $E_r(B) = - \sum_{i=1}^m \frac{1}{n} \log_2 \frac{1}{|[X_i]_{R_B^*}|}$

Утверждение 3.9. $0 \leq E_r(B) \leq \log_2 n$

Более того, если $R_B^* = \Delta$, то $E_r(B)$ достигает минимального значения 0; если $R_B^* = \omega$, то $E_r(B)$ достигает максимального значения $\log_2 n$.

Определение 3.6. Объем информации подсистемы (U, V) определяется следующим образом:

$$E(B) = \sum_{i=1}^m \frac{|X_i|}{n} \frac{|U - X_i|}{n}, \text{ где } U \setminus R_B^* = \{X_1, X_2, \dots, X_n\}$$

Утверждение 3.10. $E(B) = \sum_{i=1}^n \frac{1}{n} \frac{\frac{\infty}{\zeta} - \frac{|[X_i]_{R_B^*}|}{n}}{\frac{\infty}{\zeta}}$

Утверждение 3.11. $G(B) + E(B) = 1$.

Был проведен численный эксперимент с четырьмя наборами данных, полученными из UCI (Хранилища баз данных машинного обучения), который показан в табл. 3.1, и проведено сравнение четырех инструментов для измерения неопределенности HIS.

Таблица 3.1

4 набора данных из UCI

N	Набор данных	Образцы	Масштабированный	Порядковый	Номинальный	Классы
1	Набор 1	368	5	2	15	2
2	Набор 2	583	5	4	1	2
3	Набор 3	155	2	4	13	2
4	Набор 4	270	1	6	6	2

Результаты эксперимента показаны на рис. 3.1. Можно предположить, что грануляция информации G и грубая энтропия E_r монотонно уменьшаются по мере увеличения отношений эквивалентности. В то же время количество информации E и информационная энтропия H монотонно увеличиваются с увеличением отношений эквивалентности. Это означает, что неопределенность нечеткого отношения уменьшается по мере увеличения отношений эквивалентности. Таким образом, грануляция информации G , грубая энтропия E_r , объем информации E и информационная энтропия H могут быть применены для измерения неопределенности HIS.

Также проведен дисперсионный, корреляционный анализ, тесты Фридмана и Бонферрони–Дана.

Разработаны алгоритмы редукции HIS, основанные на грануляции информации и информационной энтропии, рис. 3.2.

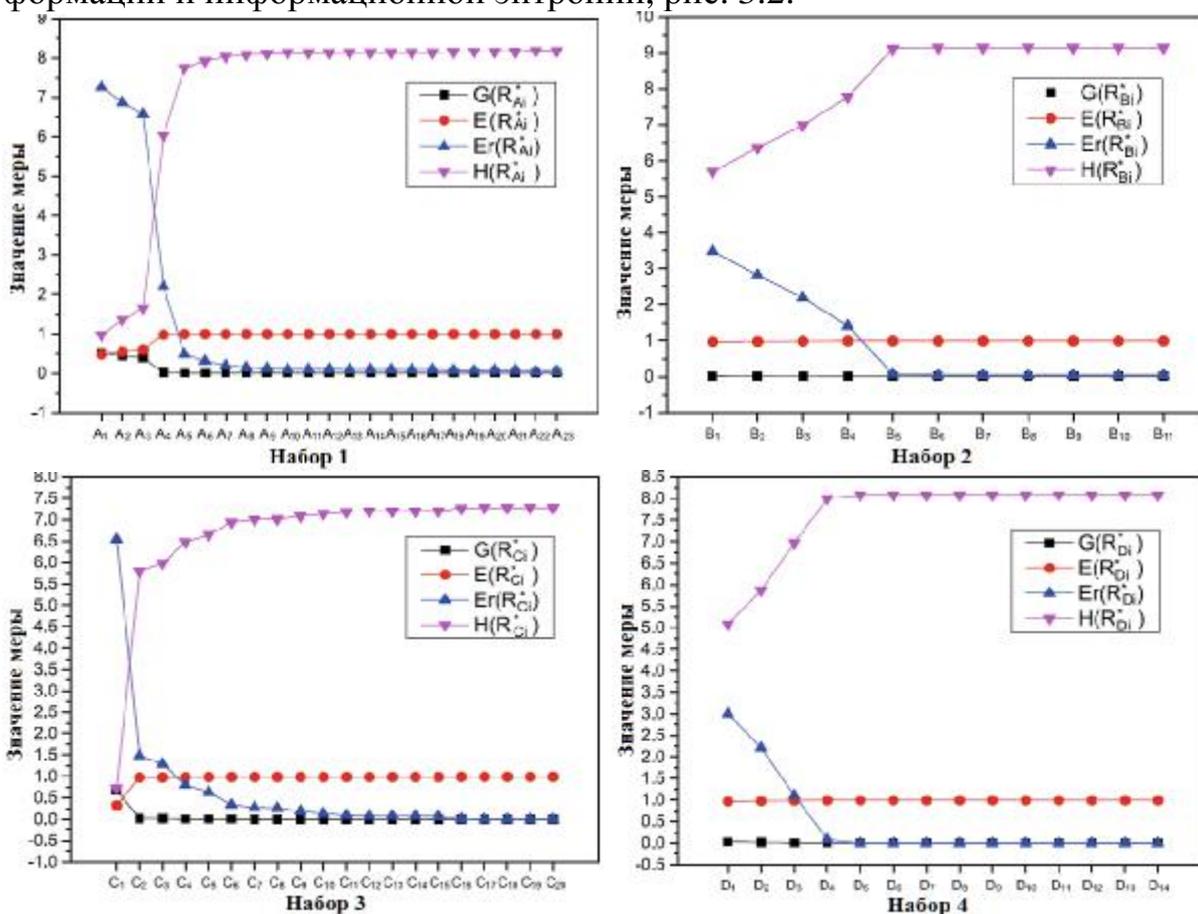


Рис. 3.1. CV-значения четырех наборов измерений

input: A HIS (U,A).
output: A reduct B
begin

```

B=A
start=1
Compute G(A)
while start do
    for каждого атрибута  $a \in B$  do
        if  $G(B - \{a\}) = G(B)$ 
        then
             $B = B - \{a\}$ 
        else
            start = 0
        end if
    end for
end while
return B

```

end

Алгоритм 3.1 редукции в HIS, основанный на грануляции информации

input: A HIS (U,A).
output: A reduct B.
begin

```

B=0
start = 1
Compute H(A)
while start do
    for каждого атрибута  $a \in A - B$  do
        if  $H(B \cup \{a\}) \neq H(A)$ 
        then
             $B = B \cup \{a\}$ 
        else
            start = 0
        end if
    end for
end while
return B

```

end

Алгоритм 3.2 редукции в HIS, основанный на информационной энтропии

Рис. 3.2. Алгоритмы редукции HIS

Алгоритм 3.1 редукции в HIS, основанный на грануляции информации, для редукции атрибута использует детализацию информации в каждом цикле. Алгоритм завершается, когда редукция любых оставшихся атрибутов не приводит к уменьшению функции оценки.

Алгоритм 3.2 редукции в HIS, основанный на информационной энтропии, в цикле получает атрибуты для объединения с текущим набором. Алгоритм завершается, когда добавление любого оставшегося атрибута не приводит к уменьшению расчетной функции.

С использованием Алгоритма 3.1 выполнено 19 поисков и каждый раз получено по одному уменьшению для каждого набора данных, указанного в табл. 3.2. На рис. 3.3 показан пример редукции для кластеризации KMeans для Набора 5 с двумя кластерами.

Таблица 3.2

Пять наборов данных из UCI

Набор данных	Образцы	Условные атрибуты				Классы
		Все	Номинальные	Порядковые	Масштабируемые	
Набор 4	299	12	5	6	1	2
Набор 5	690	14	6	5	3	2
Набор 6	690	15	9	3	3	2
Набор 7	270	13	6	6	1	2
Набор 8	303	13	8	4	1	5

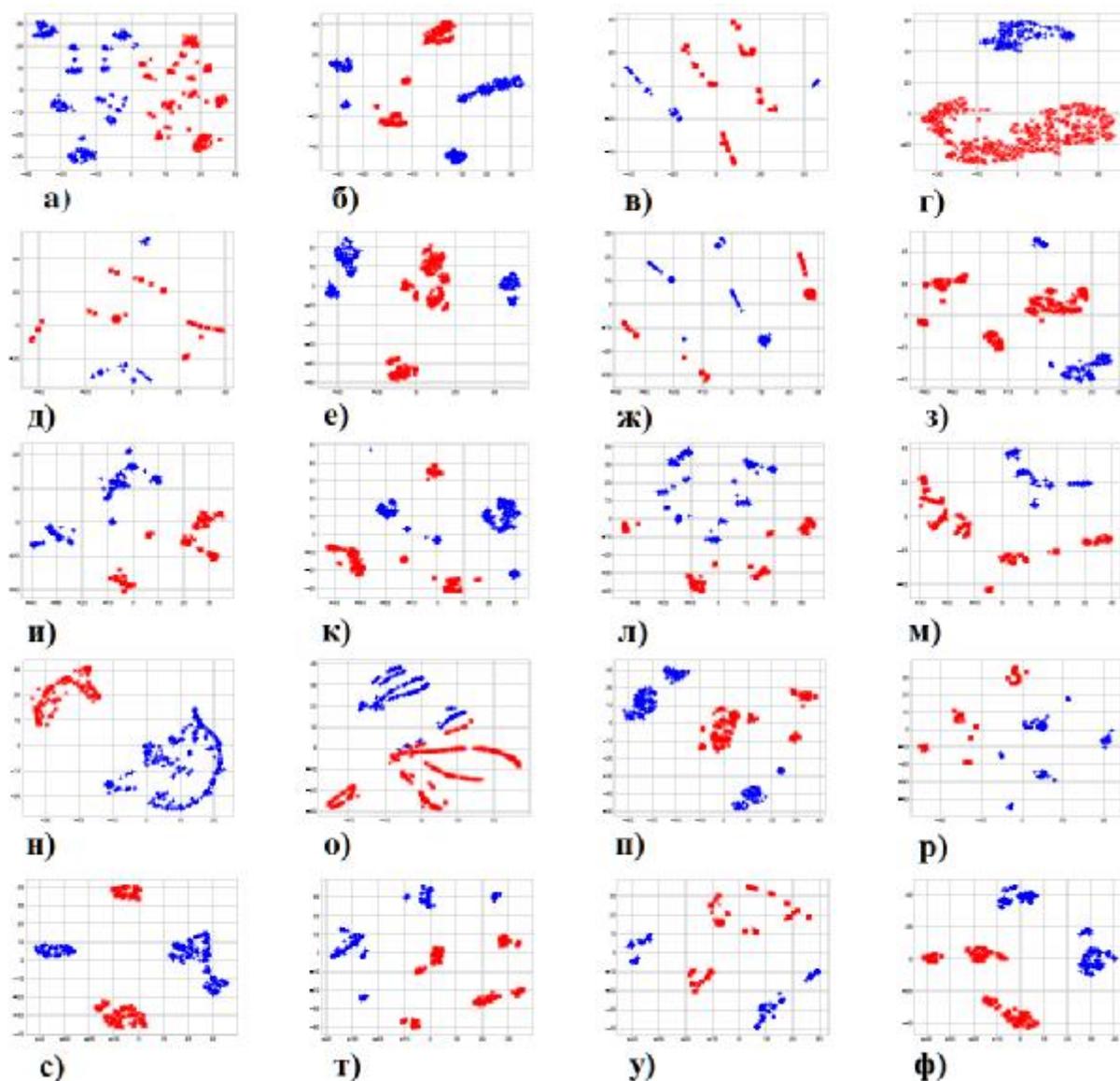


Рис. 3.3 Кластеризация с использованием KMeans на Наборе 5 с двумя кластерами: а – исходные данные; б) редукция A_1 ; в) редукция A_2 ; г) редукция A_3 ; д) редукция A_4 ; е) редукция A_5 ; ж) редукция A_6 ; з) редукция A_7 ; и) редукция A_8 ; к) редукция A_9 ; л) редукция A_{10} ; м) редукция A_{11} ; н) редукция A_{12} ; о) редукция A_3 ; п) редукция A_{14} ; р) редукция A_{15} ; с) редукция A_{16} ; т) редукция A_{17} ; у) редукция A_{18} ; ф) редукция A_{19}

Разработана архитектура программной системы управления гетерогенной информационной системой, отличающаяся использованием отношения эквивалентности на множестве объектов для измерения неопределенности системы, и обеспечивающая редукцию многомерных анализируемых атрибутов на основе грануляции информации и информационной энтропии. Архитектура представлена на рис. 3.4.

Таким образом, предложена архитектура программной системы управления гетерогенной информационной системой, отличающаяся использованием отношения эквивалентности на множестве объектов для измерения неопределенности системы, и обеспечивающая редукцию

многомерных анализируемых атрибутов на основе грануляции информации и информационной энтропии.



Рис. 3.4. Архитектура программной системы управления гетерогенной информационной системой

В главе 4 проанализированы особенности управления гетерогенными данными.

В новом подходе, основанном на пуле, процесс маркировки выполняется пакетами, выбранными из пула немаркированных данных; после маркировки каждого пакета алгоритм обучается с использованием этих пакетов; и этот процесс повторяется с набором новых образцов до тех пор, пока эффективность обучения не улучшится.

Этап предварительной обработки включает в себя методы удаления избыточных и незначительных данных для минимизации объема объекта.

Этап извлечения данных использует методы являются BOW, TF-IDF и word2vec.

Был создан групповой классификатор и оценен с использованием всех созданных показателей. Точность (4.1), погрешность (4.2), реактивность (4.3) и оценка F1 (4.4) - оценочные показатели:

$$\text{Accuracy} = (TP + TN) / (TP + FP + FN + TN) \quad (4.1)$$

$$\text{Precision} = TP / (TP + FP) \quad (4.2)$$

$$\text{Recall} = TP / (TP + FN) \quad (4.3)$$

$$F1 = 2 \times \text{Precision} \times \text{Recall} / (\text{Precision} + \text{Recall}) \quad (4.4)$$

TP, FN, FP и TN определяются количеством реальных положительных случаев, количеством ложноотрицательных случаев и количеством истинно отрицательных случаев.

Используется выборка с неопределенностью, при которой из массива немаркированных данных отбираются наименее достоверные образцы. Затем эти образцы обрабатываются специалистом, и этот процесс повторяется до тех пор, пока не будут помечены все партии. После каждой партии точность проверяется до тех пор, пока не будет достигнута требуемая степень точности. Затем помеченные образцы данных классифицируются. Этот процесс повторяется для маркировки большего количества немаркированных образцов до тех пор, пока точность модели не повысится до приемлемого уровня.

Активное обучение на основе пула данных является наиболее распространенной формой активного обучения, процесс обучения и маркировки выполняется пакетами, извлекаемыми из пула немаркированных данных. Как показано на рис. 4.1, алгоритм обучения обучается с использованием помеченных образцов после маркировки каждой партии и выбора новых образцов для обучения, что может повысить эффективность обучения; алгоритм может быть записан так, как показано на рис. 4.2. Но данные передаются алгоритму в виде потока со всеми обучающими выборками при активном обучении на основе потока. Каждый случай независимо передается алгоритму для рассмотрения. Этот пример должен быть помечен или не помечен непосредственно алгоритмом. Некоторые примеры обучения в этом пуле помечаются экспертом-человеком, и непосредственно перед отображением следующего примера алгоритм получает метку.



Рис. 4.1. Процесс активного обучения

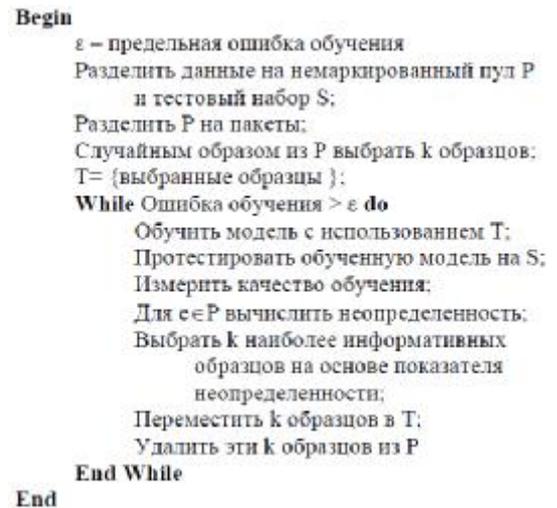


Рис. 4.2. Алгоритм 1 активного обучения на основе ансамблевой модели

Качество используемых данных оказывает значительное влияние на результаты задачи классификации. Процесс маркировки данных является ключевым препятствием для классификации, особенно при нынешней доступности ресурсов данных. Следовательно, неоднородность данных стала новой проблемой для обработки данных, выбор наиболее полезных данных зависит

от используемой меры неопределенности. Алгоритм активного обучения в выборке на основе пула выбирает экземпляры, которые вносят свой вклад в увеличивающуюся обучающую выборку, наиболее полезными выборками из которой являются те, которые наименее достоверны.

В эксперименте программирование было реализовано в среде Anaconda, которая представляет собой универсальную среду разработки на языке Python. Эксперименты проводились на устройстве со следующими характеристиками: платформа Windows 10, 8 ГБ оперативной памяти, процессор Core i5. Оценка методов классификации проведена, используя функции BOW, TF-IDF и word2vec.

В табл. 4.1 приведены результаты эксперимента по оценке точности традиционного классификатора ML и проведено сравнение выходных данных каждого классификатора с результатами ансамблевой модели, использованной в наборе данных HCR с помощью TF-IDF при разработке модели.

Таблица 4.1

Сравнительный анализ результатов традиционных классификаторов и ансамблевой модели: набор данных HCR с TF-IDF

Метод	T	Precision	Recall	F1-score
Логистическая регрессия	0.78	0.79	0.79	0.79
Случайный лес	0.80	0.79	0.79	0.78
SVM	0.79	0.79	0.80	0.79
Наивный Байес	0.80	0.78	0.79	0.78
Ансамблевая модель	0.85	0.84	0.85	0.85

Результаты эксперимента показывают, что ансамблевый классификатор достиг наилучшей производительности среди всех других классификаторов с точностью 85%, погрешностью 0,84%, повторением 85% и F1-мерой 85%, используя методы извлечения признаков (TF-IDF).

Таким образом, разработан алгоритм выбора наилучшей модели машинного обучения для анализа гетерогенных данных, отличающийся применением активного обучения и ансамблевых методов для классификаторов при решении проблем анализа разнородных данных, обеспечивающий сокращение объемов данных и повышение точности анализа совокупности данных.

Далее в работе проведена оценка неопределенностей нулевого значения гетерогенной базы данных на основе искусственного интеллекта.

Из-за сложности объективного мира потеря гетерогенной информации и неопределенность являются обычным явлением. В качестве инструмента для отображения реального мира база данных использует нулевые значения (null), чтобы выразить проблему отсутствия информации. Для решения проблемы нулевого значения в базе данных с неопределенностью предлагается алгоритм оценки нулевого значения на основе искусственного интеллекта. Сначала анализируются характеристики неопределенной базы данных, затем строится модель поиска потерянной информации, а оценка пустого значения

базы данных завершается выбором признаков и преобразованием данных, кластеризацией искусственного интеллекта, вычислением степени влияния, оценкой шага пустого значения и другими методами. Наконец, он анализирует временную сложность алгоритма и устраняет проблему низкого эффекта оценки традиционных алгоритмов. Результаты показывают, что предложенный алгоритм обладает более высокой точностью, чем традиционный алгоритм.

Вопрос точной оценки значения null в процессе предварительной обработки данных обычно решается одним из следующих способов:

1. удаление записей с нулевыми значениями;
2. замена нулевого значения постоянным значением;
3. выбор среднего значения вместо нулевого значения в диапазоне значений null;
4. в диапазоне значений null вместо нулевого значения используется случайное значение;
5. статистическая функция распределения исходных данных, а затем в соответствии с функцией распределения сгенерировать значение замены нулевого значения.

Однако вышеприведенные методы не могут идеально решить все проблемы с нулевыми значениями, процесс вычисления сложен, а тенденция кластеризации исходных данных игнорируется, и эффект оценки нулевых значений может быть получен не очень хорошо.

Предлагается основанный на искусственном интеллекте алгоритм оценки нулевых значений для неопределенных баз данных. Этот метод использует принцип ошибки для определения порядка оценки нулевых значений для каждого столбца. С помощью интеллектуального анализа данных определяется набор атрибутов, связанный с оцененными атрибутами. Исходные данные разбиваются на нечеткие кластеры с помощью метода искусственного интеллекта, а нулевое значение оценивается методом линейной регрессии внутри каждого кластера.

Чтобы лучше выразить атрибут времени данных, данные в базе данных представлены в виде графика временных данных.

Граф временных данных представлен формулой $G=(V_t, E_t)$, где, V_t представляет набор временных узлов, а E_t представляет набор временных ребер.

Временным узлом является v_t , выраженный как $v_t=[v, (ts_{v_t}, te_{v_t})]$, где v представляет идентификацию временного узла, $[ts_{v_t}, te_{v_t}]$ представляет полуоткрытый временной интервал, а E - время действия данных.

Временной интервал e_t выражается как $e_t=[u, v_t, (ts', te')]$, $[ts', te']$ - эффективное время поиска.

Временные данные показаны на рис. 4.3 Как показано на рис. 4.3, информация имеет время действия и время транзакции. В процессе поиска потерянных данных в основном учитывается время действия информации.

Чтобы обеспечить извлечение как можно большего числа узлов с ключевыми словами и обеспечить тесную связь результатов поиска с информа-

цией о потерянных данных в запросе, рассчитаем вес структуры узлов.

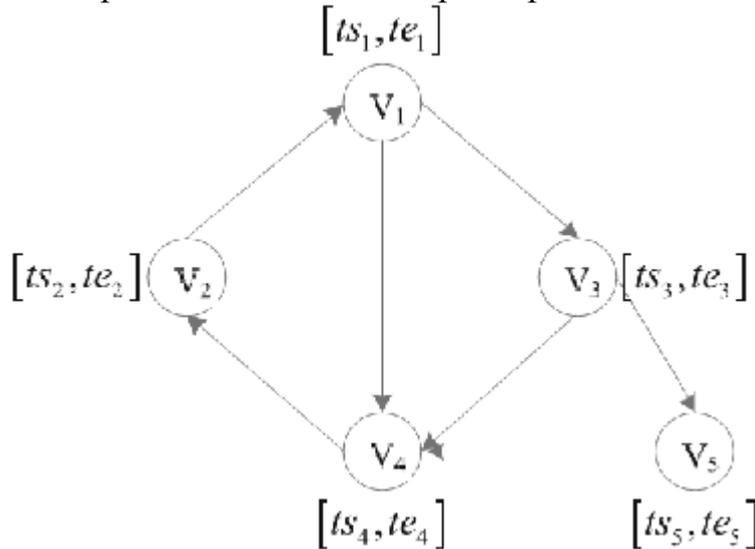


Рис. 4.3. Граф временных данных

Вес структуры узла представляет важность узла в графе временных данных, который вводится в формулу расчета веса ребра, и формула расчета веса ребра получается следующим образом:

$$W(Q, e_t) = \frac{1}{IR_{(k,u)} + IR_{(k,v)}} \cdot W_e(u, v) \quad (4.5)$$

где $W_e(u, v)$ - вес структуры узла.

Вес временного ребра устанавливается в соответствии со временем запроса потерянных данных, и формула его расчета такова:

$$W(Q, e_t) = 1 - \frac{|I_c \ominus I_t|}{I_c} \quad (4.6)$$

где I_c представляет время запроса потерянных данных, I_t - эффективное время временного интервала.

Для более точной оценки null в реляционной базе данных на основе модели поиска потерянных данных и сходства потерянных данных предлагается метод оценки пустого значения в базе данных с неопределенностью, основанный на искусственном интеллекте.

Шаг 1: Выбор объекта и преобразование данных.

1.1. Выбор объекта, алгоритм сокращения атрибутов, основанный на грубом наборе, используется для уменьшения атрибутов исходной таблицы данных и получения набора ключевых атрибутов после сокращения.

1.2. Преобразование данных относится к предварительной обработке данных, что упрощает использование формы данных. Сначала семантические атрибуты естественного языка нумеруются, чтобы их было удобно использовать для интеллектуального анализа данных. Затем используется формула нечеткого числа для нормализации числовой информации и упрощения вычислений.

Шаг 2: Кластеризация с искусственным интеллектом. Наборы ненулевых атрибутов, связанные с атрибутами с нулевыми значениями, полученными

ми на шаге 1, используются для кластеризации. При объединении похожих данных разные данные разбиваются на разные кластеры. На рис. 4.4 показана совместимость объектов из разных наборов атрибутов.

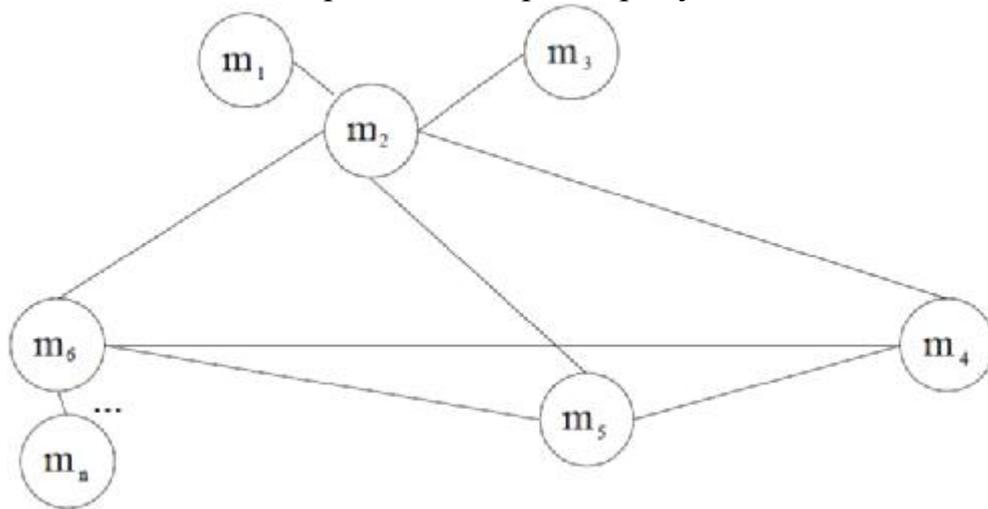


Рис. 4.4. Соотношение совместимости объектов в разных наборах атрибутов

Как показано на рис. 4.4, учитывая, что разные атрибуты имеют разный вес влияния на столбцы с нулевыми значениями, вводятся соответствующие веса:

$$w = \frac{r^2 - W(Q, e_t) \phi}{\sum_{k=1}^m r_k^2} \quad (4.7)$$

где m - количество атрибутов в наборе непустых атрибутов, относящихся к атрибутам с нулевыми значениями; r - коэффициенты корреляции атрибутов с нулевыми значениями; w - отношение коэффициентов корреляции атрибутов с нулевыми значениями и сумма коэффициентов корреляции всех связанных атрибутов

После кластеризации с использованием искусственного интеллекта будет получен центр кластеризации.

Шаг 3: Вычисление влияния. После кластеризации данных в несколько кластеров для каждого кластера влияние различных независимых переменных на зависимые переменные различно. Сначала коэффициент нечеткой корреляции используется для представления степени корреляции между атрибутами, затем определяется коэффициент независимой переменной и, наконец, получается степень влияния атрибутов. Формула для расчета степени корреляции:

$$Z_{a,b} = \frac{\sum_{i=1}^n (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_{i=1}^n (a_i - \bar{a})^2 \sum_{i=1}^n (b_i - \bar{b})^2}} \quad (4.8)$$

где \bar{a} и \bar{b} - выборочное среднее значение a , b нечеткого множества.

Формула для определения коэффициента независимой переменной:

$$\text{COD} = \pm \frac{r^2}{\sum_{k=1}^m r_1^2} \quad (4.9)$$

Шаг 4: Оценка нулевых значений.

Проведен эксперимент по оценке качества алгоритма. Окружение эксперимента: операционная система - Windows 7, процессор Intel Core i7-3770 с тактовой частотой 3,40 ГГц, 8 ГБ оперативной памяти, язык C+ на платформе разработки Microsoft Visual Studio 2012.

По результатам эксперимента алгоритм оценки нулевого значения в базе данных с неопределенностью, основанный на искусственном интеллекте, обеспечивает более эффективную и точную оценку нулевых значений в среднем на 7.6%.

Таким образом, представлен алгоритм идентификации нулевых значений в гетерогенных базах данных, отличающийся предварительной классификацией исходных данных на основе взвешенных значений и обеспечивающий более эффективную и точную оценку нулевых значений в среднем на 7.6%.

Заключение

В процессе выполнения диссертационного исследования получены следующие основные результаты:

1. Проведен анализ проблем управления гетерогенными данными информационных систем с многомерными атрибутами на основе мягкой максиминной оценки и активного обучения.

2. Разработана мягкая максиминная оценка для гетерогенных данных, содержащих уникальные вариационные компоненты, обеспечивающая сохранение статистических свойств и лучшую вычислительную эффективность

3. Предложена архитектура гетерогенной программной системы, обеспечивающая эффективную редукцию многомерных анализируемых атрибутов.

4. Создан алгоритм выбора наилучшей модели машинного обучения для анализа гетерогенных данных, обеспечивающий сокращение объемов данных и повышение точности анализа совокупности данных.

5. Разработан алгоритм идентификации нулевых значений в гетерогенных базах данных, обеспечивающий более эффективную и точную оценку нулевых значений в среднем на 7.6%.

6. Элементы программного обеспечения зарегистрированы в ФИПС.

Рекомендации и перспективы дальнейшей разработки темы

1. Результаты исследования рекомендуются к применению в задачах управления гетерогенными данными информационных систем с многомерными атрибутами.

2. Дальнейшая разработка темы будет направлена на практическую реализацию теоретических и алгоритмических результатов, интеграцию в наиболее распространенные распределенные системы. Развитие результатов

будет направлено на улучшение модифицируемости и реконфигурируемости программных систем.

Основные результаты диссертации опубликованы в следующих работах:

Публикации в изданиях списка ВАК

1. Атласов Д.И., Кравец О.Я., Пашкевич А.С. Управление специальной информацией в беспроводной сети на основе гетерогенной последовательности данных// Системы управления и информационные технологии, №1(91), 2023. – С. 49-55.

2. Атласов Д.И., Красновский Е.Е., Сараев П.В. Пути интерполяции мягкой максиминной оценки для гетерогенных данных// Системы управления и информационные технологии, №4(94), 2023. С. 27-30.

3. Сотников Д.В., Атласов Д.И., Кравец О.Я., Красновский Е.Е. Исследование метода оптимизации данных для эксплуатации и сопровождения базы знаний программного обеспечения на основе облачных вычислений// Системы управления и информационные технологии, №2(100), 2025. С. 37-42.

4. Атласов Д.И., Сотников Д.В., Кравец О.Я., Красновский Е.Е. Оценка неопределенности нулевых значений базы данных на основе искусственного интеллекта// Системы управления и информационные технологии, №2.1(100), 2025. С. 4-11.

5. Атласов Д.И., Васми И., Коптелова А.С., Кочегаров А.В. Оценка и оптимизация систем с гетерогенными данными с учетом показателей эффективности на основе интегрированного алгоритма// Моделирование, оптимизация и информационные технологии. 2025; 13(3). URL: <https://moitvivr.ru/ru/journal/pdf?id=2014>. DOI: 10.26102/2310-6018/2025.50.3.025.

Публикация в издании, входящем в базу данных WoS

6. Kravets O.Ja., Atlasov D.I. et al. Designing the architecture of a distributed system for information monitoring of IoT and IIoT infrastructures traffic// International Journal on Information Technologies and Security, vol. 16, no. 1, 2024, pp. 49-56. <https://doi.org/10.59035/VTBI7690>.

Свидетельство о государственной регистрации программы для ЭВМ

7. Атласов Д.И., Сотников Д.В., Васми Ихаб А Васми, Хуссейн Али Иед, Линкина А.В. Типовой интерфейс облачных вычислений. Свидетельство о регистрации программы для ЭВМ № 2025681822 от 18.08.2025. - М.: Роспатент, 2025.

Статьи и материалы конференций

8. Atlasov D.I., Kravets O.Ja. To the formulation of the problem of extracting a common signal from heterogeneous data of heterogeneous information systems// Modern informatization problems in simulation and social technologies (MIP-2023'SCT): Proceedings of the XXVIII-th International Open Science Conference (Yelm, WA, USA, January 2023). - Yelm, WA, USA: Science Book Publishing House, 2023. – Pp. 8-13.

9. Атласов Д.И., Кравец О.Я. Идентификация параметров функциональной зависимости прироста ресурсоотдачи процесса решения задач управления информационно-коммуникационными системами// Информационные технологии моделирования и управления, №2(132), 2023. – С. 110-116.

10. Atlasov D.I., Kravets O.Ja. Computational features of soft maximin estimation interpolation for heterogeneous data// Modern informatization problems in the technological and telecommunication systems analysis and synthesis (MIP-2024'AS): Proceedings of the XXIX-th International Open Science Conference. - Yelm, WA, USA: Science

Book Publishing House, 2024. Pp. 120-124.

11. Атласов Д.И., Кравец О. Я. Извлечение надежного сигнала из гетерогенных данных// Современные инновации, системы и технологии - Modern Innovations, Systems and Technologies, 2024, 4(1), 0122–0132. <https://doi.org/10.47813/2782-2818-2024-4-1-0122-0132>.

12. Атласов Д.И., Кравец О.Я. Эффективный метод получения необходимой информации из набора данных, представляющих собой совокупность неоднородных групп// Оптимизация и моделирование в автоматизированных системах: тр. Междунар. молодежной научной школы. – Воронеж: ВГТУ, 2023. С. 66-70.

13. Атласов Д.И., Кравец О.Я. Разработка системы обнаружения аномалий на основе изолированного леса// Интеллектуальные информационные системы: тр. Междунар. НПК, посв. 40-летию кафедры САПРИС. - Воронеж, 2024. – с. 180-184.

14. Атласов Д.И. Нечеткая энтропия для разнородных данных в гетерогенных информационных системах// Сб. тр. VI Всеросс. НПК «Информационные технологии в экономике и управлении». – Махачкала, 2024. С. 65-70.

15. Атласов Д.И., Кравец О.Я. Экспериментальное исследование по устранению неопределенности данных в гетерогенных информационных системах// Экономика и менеджмент систем управления, №1(55), 2025. – С. 22-31.

16. Atlasov D.I., Kravets O.Ja. Theoretical foundations for measuring the uncertainty of heterogeneous data// Modern informatization problems in simulation and social technologies (MIP-2025'SCT): Proc. of the XXX-th Int. Open Science Conf. - Yelm, WA, USA: Science Book Publishing House, 2025. – pp. 11-23.

17. Атласов Д.И., Кравец О.Я. Комплексный подход к обработке гетерогенных данных с активным обучением// Информационные технологии моделирования и управления, №4(142), 2025. – С. 304-312.

18. Атласов Д.И., Кравец О.Я. Экспериментальное исследование по устранению неопределенности данных в гетерогенных информационных системах// Экономика и менеджмент систем управления, №1(55), 2025. – С. 22-31.

19. Atlasov D.I. An experimental study of an integrated approach to heterogeneous data processing with active learning// Modern informatization problems in simulation and social technologies (MIP-2026'SCT): Proc. of the XXXI-th Int. Open Science Conf. - Yelm, WA, USA: Science Book Publishing House, 2026. – pp. 20-28.

Подписано в печать 13.02.2026.

Формат 60x84/16. Бумага для множительных аппаратов.

Усл. печ. л. 1,0. Тираж 80 экз. Заказ №74.

ФГБОУ ВО «Воронежский государственный технический университет»

394026 Воронеж, Московский просп., 14