

УДК 004.89

АЛГОРИТМЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ В РАСПРЕДЕЛЁННЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ ПРИ ОГРАНИЧЕНИЯХ АСИНХРОННОГО ОБМЕНА

Яницкая Т. С.¹, Усманов Р. И.¹¹Поволжский государственный университет сервиса, Тольятти, Россия

Статья поступила 30.11.2025, принята к публикации 30.12.2025. Опубликовано онлайн.

Аннотация. Рассматривается задача интеллектуального анализа данных в распределённых информационных системах при асинхронном обмене между узлами. Анализ выполняется в условиях отсутствия глобальной синхронизации и неполной наблюдаемости состояний. Система описывается ориентированным графом с потоками данных $D_i(t)$ и задержками передачи $t_d > 0$. Предлагается алгоритмический подход, основанный на локальной обработке и последующей агрегации результатов при ограничениях на согласованность данных. Формализуется критерий качества анализа $J = \sum_{i=1}^n \phi(S_i, Y_i)$ с учётом рассинхронизации состояний.

Рассматриваются исключительные ситуации, связанные с потерями сообщений и неупорядоченной доставкой. Приводится оценка вычислительной сложности и устойчивости алгоритмов к росту степени распределённости. Результаты ориентированы на применение в распределённых аналитических и управляющих системах.

Ключевые слова: интеллектуальный анализ данных, распределённые информационные системы, асинхронный обмен, алгоритмы анализа, задержки передачи данных, агрегация результатов, вычислительная сложность

ALGORITHMS FOR INTELLIGENT DATA ANALYSIS IN DISTRIBUTED INFORMATION SYSTEMS UNDER ASYNCHRONOUS COMMUNICATION CONSTRAINTS

Ianitskaia T. S.¹, Usmanov R. I.¹¹Volga Region State University of Service, Tolyatti, Russia

Abstract. The problem of intelligent data analysis in distributed information systems under asynchronous exchange between nodes is considered. The analysis is performed under conditions of the absence of global synchronization and partial observability of states. The system is described as a directed graph with data streams $D_i(t)$ and transmission delays $t_d > 0$. An algorithmic approach is proposed, based on local processing and subsequent aggregation of results under constraints on data

consistency. A quality criterion of analysis $J = \sum_{i=1}^n \phi(S_i, Y_i)$ is formalized, taking into account state desynchronization.

Exceptional situations related to message loss and out-of-order delivery are examined. The computational complexity and robustness of the algorithms with respect to increasing system distribution are evaluated. The results are intended for application in distributed analytical and control systems.

Keywords: intelligent data analysis, distributed information systems, asynchronous communication, analysis algorithms, data transmission delays, result aggregation, computational complexity

Введение

Распределённые информационные системы применяются для обработки данных в условиях пространственной разнесённости источников, неоднородности вычислительных ресурсов и отсутствия глобального управляющего центра. В таких системах обмен данными между узлами осуществляется асинхронно, с переменными задержками передачи и возможными потерями сообщений. Указанные свойства приводят к частичной наблюдаемости состояния системы и исключают возможность формирования согласованного глобального состояния в фиксированный момент времени.

Интеллектуальный анализ данных (ИАД) в распределённой среде принципиально отличается от централизованных систем. Алгоритмы, предполагающие наличие полной и синхронизированной выборки, оказываются неприменимыми или демонстрируют неконтролируемый рост вычислительной сложности. Основная причина заключается в невозможности гарантировать согласованность входных данных по времени и полноте. В реальных распределённых системах данные поступают фрагментарно, с запаздываниями и в неупорядоченном виде.

Актуальной задачей является разработка алгоритмов интеллектуального анализа данных, устойчивых к асинхронному обмену, задержкам передачи и потерям сообщений. Такие алгоритмы должны опираться на локальную обработку информации и обеспечивать корректную агрегацию частичных результатов без предположений о глобальной синхронизации. Настоящая работа

направлена на формализацию математической модели распределённой информационной системы и разработку алгоритмического подхода к интеллектуальному анализу данных в условиях асинхронного обмена.

Методы

Распределённая информационная система моделируется в виде ориентированного графа $G = (V, E)$, где $V = v_1, \dots, v_n$ – множество вычислительных узлов, $E \subseteq V \times V$ – множество каналов передачи данных. Каждому ребру $e_{ij} \in E$ сопоставляется задержка передачи $t_{ij} > 0$ и вероятность потери сообщения $p_{ij} \in [0, 1]$. Глобальная синхронизация между узлами отсутствует [1].

На каждом узле v_i формируется поток данных $D_i(t)$, определяемый как последовательность наблюдений, поступающих в дискретные или непрерывные моменты времени t . Поток может содержать пропуски, а временные метки наблюдений не предполагаются согласованными между узлами. Локальное состояние узла обозначается как $S_i(t)$ и формируется на основе доступной на узле информации.

Рассмотрим динамику локального состояния узла v_i . Данный показатель описывается следующим соотношением: $S_i(t) = F_i(D_i(t), M_i(t))$, где: $S_i(t)$ – локальное состояние узла v_i в момент времени t , $F_i(\cdot)$ – локальный алгоритм обработки данных, который зависит от локальных данных $D_i(t)$ и совокупности сообщений $M_i(t)$, полученных от соседних узлов в момент времени t . Введённых нами обозначениях $F_i(\cdot)$ и ниже $H_i(\cdot)$, $\Phi(\cdot)$ – точкой в скобках мы указываем на общее свойство, не уточняя на данном этапе конкретные значения. Это абстрактное выражение функций.

Алгоритм $F_i(\cdot)$ представляет собой функцию, которая вычисляет состояние узла на основе данных, поступающих от других узлов. Таким образом, состояние узла зависит как от локальных данных, так и от внешней информации, получаемой через сообщения.

Сообщения $M_i(t)$ могут содержать агрегированные данные (обобщённые), вычисленные другими узлами. Эти сообщения поступают с задержками t_{ji} , которые определяются временем, необходимым для передачи данных по каналам связи между узлами.

На этом этапе необходимо пояснить: узлы обозначены как v_i . Нам требуется показать связь между узлами. Узлы получающие сообщения мы описываем как v_i , узлы передающие сообщения – это v_j . Соответственно t_{ji} – это время, которое требуется для передачи сообщения от узла v_j к узлу v_i .

Важным элементом алгоритма является функция отклонения, которая используется для измерения разницы между состоянием узла $S_i(t)$ и его эталонным состоянием Y_i . Такая функция может быть разной в зависимости от выбранной метрики для оценки качества локального состояния.

Здесь может быть рассмотрена либо квадратичная функция отклонения, либо абсолютная функция отклонения. Выбор зависит от того, чувствителен ли результат к отклонениям. Квадратичная функция отклонения описывается как $F(S_i, Y_i) = (S_i - Y_i)^2$, абсолютная функция отклонения описывается уравнением вида: $F(S_i, Y_i) = |S_i - Y_i|$.

Отклонение характеризует разницу между состоянием узла и эталонным значением или «нормальным» для узлов. В зависимости от контекста задачи, можно использовать более точные или более устойчивые к выбросам (outliers) метрики. Outliers – это существенные, аномальные

отклонения данных, которые могут в значительной степени исказить результаты анализа. Абсолютная функция отклонения менее чувствительна к выбросам. И квадратичная функция для области анализа распределённых систем является, по мнению авторов, более предпочтительной [2].

Чтобы убрать зависимости от глобального состояния и свести к минимуму влияния асинхронности обмена данными используется принцип асинхронной агрегации. Такой подход позволяет каждому узлу работать с локальной информацией. При этом нет необходимости использовать данные по глобальной синхронизации состояния всей системы.

Каждый узел v_i периодически формирует агрегированный вектор $A_i(t) = H_i(S_i(t))$, где $H_i(\cdot)$ – оператор сжатия локального состояния. Оператор $H_i(\cdot)$ выполняет агрегацию, т.е. состояние узла $S_i(t)$ может быть сложным, многомерным и оператор объединяет это состояние (сжимает) в один показатель $A_i(t)$, который может быть передан другим узлам.

В распределённых системах такой подход позволяет сократить количество передаваемой информации и снизить нагрузку на сетевые каналы передачи [3].

Алгоритм агрегации может использовать различные методы сжатия, такие как медиану (Median, является более устойчива к outliers или ранг (Rank), усреднение (Average), или другие подходы (взвешенная агрегация; дисперсия, и стандартное отклонение, и т.д.).

Точное поведение оператора агрегации зависит от специфики задачи. В рамках данного исследования оператор $H_i(\cdot)$ выполняет агрегацию с целью получения обобщенного признака для дальнейшего обмена информацией между узлами. Вектор $A_i(t)$ передаётся по рёбрам графа G соседним узлам.

Это позволяет узлам обмениваться агрегированными данными, а не исходными потоками. Исходные потоки данных $D_i(t)$ между узлами не передаются, что минимизирует зависимость от полной синхронизации состояния системы.

Таким образом, за счёт асинхронной агрегации, система может функционировать без необходимости в глобальной синхронизации состояния всех узлов. Это делает её более устойчивой к задержкам, потере сообщений [4].

Качество интеллектуального анализа оценивается функционалом $J = \sum_{i=1}^n \phi(S_i, Y_i)$, где Y_i – эталонное или ожидаемое значение состояния узла v_i , а $\phi(\cdot)$ – функция отклонения. Функционал J определяется на множестве асинхронных состояний и не предполагает одновременного вычисления для всех узлов.

Для учёта устаревших сообщений вводится порог задержки Δt_{max} . Сообщения, поступившие с задержкой, превышающей Δt_{max} , учитываются с пониженным весом $\alpha \in (0,1)$ либо исключаются из агрегации. Вес сообщения определяется функцией $\omega(\tau)$, где τ – фактическая задержка доставки.

Проверка алгоритмов выполняется методом имитационного моделирования. Задержки t_{ij} , вероятности потерь p_{ij} и интенсивности потоков $D_i(t)$ задаются параметрически. Рассматриваются сценарии сетевых разделов (т.е. сценарии при разрывах каналов, которые рамках распределённых информационных систем рассматриваются как нормальные состояния деградации, а не как исключительные ситуации), временной недоступности узлов и неупорядоченной доставки сообщений.

Результаты

Исследование поведения алгоритмов начинается с анализа локальной обработки данных. При асинхронном поступлении наблюдений локальные состояния $S_i(t)$

формируются на неполной информации. Отсутствие глобального состояния не приводит к остановке алгоритма, так как вычисления зависят только от локально доступных данных и агрегированных сообщений соседних узлов.

В ходе моделирования установлено, что передача агрегированных векторов $A_i(t)$, вместо исходных потоков $D_i(t)$, существенно снижает сетевую нагрузку. При фиксированном числе узлов n , объём передаваемых данных растёт линейно по числу рёбер графа G . Это позволяет контролировать коммуникационную сложность при масштабировании системы.

Анализ влияния задержек передачи показал, что при увеличении средних значений t_{ij} функционал качества J изменяется непрерывно и не демонстрирует значительных оклонов. Основное влияние оказывает не абсолютная величина задержки, а её вариативность. При высокой дисперсии задержек возрастает роль механизма взвешивания сообщений по времени доставки.

При увеличении числа узлов n наблюдается умеренное снижение точности анализа, выражаемое ростом значения J . Данный эффект связан с накоплением локальных погрешностей и увеличением времени распространения агрегированных состояний по графу. Однако при использовании адаптивных весов $\omega(\tau)$ снижение качества остаётся ограниченным.

В сценариях потери сообщений алгоритмы сохраняют работоспособность. При вероятностях потерь p_{ij} до заданного порога система демонстрирует устойчивое поведение. Локальные состояния компенсируют недостающую информацию за счёт собственных данных и сообщений от альтернативных маршрутов в графе G .

Вычислительная сложность локальной обработки определяется как $O(f(|D_i|))$, где $|D_i|$ – объём данных, обработанных узлом за интервал времени. Агрегация выполняется за время, линейное по числу входящих сообщений. Эксперименты подтверждают отсутствие экспоненциального роста вычислительных затрат при увеличении числа узлов.

Дополнительно проанализированы режимы кратковременной недоступности узлов. При исключении узла v_k из взаимодействия на интервале $[t_1, t_2]$ система переходит в новое асинхронное состояние. После восстановления связи агрегированные показатели постепенно интегрируются без необходимости пересчёта всей истории данных.

Передача агрегированных характеристик вместо исходных потоков данных значительно уменьшает объём передаваемой информации, что помогает минимизировать влияние потерь сообщений и задержек в передаче. Для учёта устаревших сообщений вводится порог задержки Δt_{max} сообщения, поступившие с задержкой, превышающей этот порог, получают пониженный вес или исключаются из агрегации. Вес сообщения, поступившего с задержкой, определяется функцией $\omega(\tau)$, где τ – фактическая задержка доставки сообщения.

Наше исследование описывает распределённую информационную систему прежде всего на уровне формальной модели взаимодействия узлов. Заданы графовая структура, задержки передачи, потери сообщений и отсутствие глобальной синхронизации. Эти элементы определяют среду функционирования алгоритмов, но сами по себе не образуют интеллектуальный анализ данных. Они фиксируют ограничения, в которых анализ выполняется, и тем самым задают рамки допустимых вычислительных процедур.

Интеллектуальный анализ данных начинается в момент, когда поток $D_i(t)$ интерпретируется не как произвольная последовательность сообщений, а как источник данных для построения аналитических представлений. Это могут быть события, записи распределённой базы данных, измерения, логи или транзакции. В этом контексте локальное состояние $S_i(t)$ перестаёт быть абстрактной переменной и приобретает конкретный смысл. Оно представляет собой набор вычисленных характеристик: статистики распределений, параметры моделей, оценки вероятностей или признаки, полученные в результате локальной обработки.

Важный момент заключается в том, что $S_i(t)$ формируется при неполной информации. Узел не обладает глобальным состоянием системы и не располагает синхронизированными данными других узлов. Следовательно, используемые алгоритмы анализа должны быть устойчивы к асинхронности, пропускам и неупорядоченной доставке данных. Это ограничение исключает применение методов, требующих согласованного доступа ко всему набору данных, но допускает использование инкрементных и онлайн-процедур.

Агрегация играет ключевую роль именно с точки зрения интеллектуального анализа. Оператор H_i применяется не к исходным данным, а к результатам анализа, полученным локально. В агрегированное значение $A_i(t)$ могут входить, например, оценки квантилей, параметры аппроксимирующих моделей, частичные вероятностные распределения или компактные описания признаков. Такой выбор не случаен. Эти объекты допускают корректное объединение при отсутствии синхронизации и сохраняют интерпретируемость при частичной потере сообщений.

Факт передачи агрегированных значений вместо исходных потоков данных определяет класс допустимых аналитических задач. Средние значения чувствительны к задержкам и выбросам. Квантили и робастные статистики ведут себя иначе. При асинхронном обмене они позволяют контролировать искажения, возникающие из-за несвоевременной доставки сообщений. По этой причине в распределённом анализе предпочтение отдается именно таким подходам (агрегаторам), а не полным выборкам.

Функция отклонения $\phi(S_i, Y_i)$ связывает аналитические результаты с целевыми ориентирами. Эталонное состояние Y_i не обязательно соответствует истинному глобальному значению. В распределённых системах оно часто задаётся как ожидаемое состояние, модельное приближение или результат предыдущей итерации анализа. В этом смысле критерий качества J оценивает не точность в классическом статистическом смысле, а устойчивость и согласованность аналитических оценок при асинхронном обновлении.

Особое значение имеют свойства алгоритмов при повторной доставке сообщений и сетевых сбоях. Если обновление локального состояния не является идемпотентным, повторное получение одного и того же агрегата приводит к накоплению ошибок. Для интеллектуального анализа это критично, поскольку искажённые оценки теряют аналитическую ценность. Поэтому допустимые процедуры обновления $S_i(t)$ должны сохранять корректность при дублировании сообщений.

Частичная наблюдаемость дополняет картину. Узел получает лишь фрагменты информации о системе. Для работы в таких условиях применяются вероятностные модели и методы оценки скрытых состояний. Они позволяют интерпретировать локальные данные как наблюдения, а агрегированные сообщения – как дополнительные ограничения. Это формирует аналитическую модель, в

которой неопределённость является не побочным эффектом, а встроенным элементом.

Таким образом, граф, задержки и агрегация не являются самоцелью. Они задают формальную среду, в которой интеллектуальный анализ данных реализуется через локальные вычисления, робастные функции агрегации и критерии качества, определённые на асинхронных состояниях. Связь с ИАД проявляется не на уровне архитектурных решений, а в выборе аналитических объектов, допустимых операций над ними и ограничений, при которых сохраняется интерпретируемость и корректность результатов.

Обсуждение

Полученные результаты согласуются с ограничениями распределённых систем, формализованными в теореме CAP (Consistency, Availability, Partition Tolerance). Отказ от строгой согласованности позволяет обеспечить доступность и устойчивость анализа при наличии сетевых разделов и асинхронного обмена. В предлагаемой модели корректность анализа определяется не мгновенным совпадением состояний, а устойчивостью функционала J во времени.

Следует отметить, что предлагаемый подход ориентирован на задачи интеллектуального анализа, допускающие приближённые и обновляемые оценки. Алгоритмы не предназначены для сценариев, требующих строгой транзакционной согласованности. Это ограничение является следствием выбранной модели асинхронного обмена и отсутствия глобальной синхронизации.

Практическая применимость разработанных алгоритмов связана с распределёнными аналитическими платформами, системами мониторинга и управления, а также с обработкой потоковых данных в условиях нестабильных сетей.

Формализация модели в виде ориентированного графа и использование локальной агрегации позволяют адаптировать алгоритмы к различным архитектурам распределённых информационных систем.

В дальнейшем развитие работы может быть направлено на уточнение функций агрегации $H_i(\cdot)$ и весовых коэффициентов $\omega(\tau)$, а также на анализ влияния динамически изменяющейся топологии графа G на свойства алгоритмов интеллектуального анализа данных.

Список литературы

1. Глушак, Е. В. Анализ распределения задач в системе облачных и туманных вычислений / Е. В. Глушак, Д. С. Ключев, В. И. Воловач // Т-Comm: Телекоммуникации и транспорт. – 2025. – Т. 19, № 7. – С. 25-33. – DOI 10.36724/2072-8735-2025-19-7-25-33.
2. Gorodetsky, V. I. Basic Trends of Decentralized Artificial Intelligence / V. I. Gorodetsky // Pattern Recognition and Image Analysis. Advances in Mathematical Theory and Applications. – 2023. – Vol. 33, No. 3. – P. 324-333. – DOI 10.1134/S105466182303015X.
3. Smagin, V. I. Robust extrapolation for systems with unknown input and interval uncertainty in system and observations / V. I. Smagin, K. S. Kim // Tomsk State University Journal of Control and Computer Science. – 2023. – No. 62. – P. 85-91. – DOI 10.17223/19988605/62/9.
4. Tsiramua, S. Logical-Probabilistic Modeling and Structural Analysis of Reconfigurable Systems Composed of Multifunctional Elements / S. Tsiramua, H. Meladze, T. Davitashvili // Pattern Recognition and Image Analysis. Advances in Mathematical Theory and Applications. – 2024. – Vol. 34, No. 1. – P. 144-157. – DOI 10.1134/S1054661824010218.
5. Rouban, A. I. The sensitivity coefficients for dynamic systems described by nonlinear

difference generalized equations with the distributed memory and characterized by generalized functionals / A. I. Rouban // Tomsk State University Journal of Control and Computer Science. – 2022. – No. 61. – P. 81-87. – DOI 10.17223/19988605/61

References

1. Glushak E.V., Klyuev D.S., Volovach V.I. Analiz raspredeleniya zadach v sisteme oblachnykh i tumannykh vychislenii. Analysis of task allocation in cloud and fog computing systems. T-Comm: Telekommunikatsii i transport. 2025, vol. 19, no. 7, pp. 25-33. DOI 10.36724/2072-8735-2025-19-7-25-33. In Russ.
2. Gorodetsky, V. I. Basic Trends of Decentralized Artificial Intelligence / V. I. Gorodetsky // Pattern Recognition and Image Analysis. Advances in Mathematical Theory and Applications. – 2023. – Vol. 33, No. 3. – P. 324-333. – DOI 10.1134/S105466182303015X.
3. Smagin, V. I. Robust extrapolation for systems with unknown input and interval uncertainty in system and observations / V. I. Smagin, K. S. Kim // Tomsk State University Journal of Control and Computer Science. – 2023. – No. 62. – P. 85-91. – DOI 10.17223/19988605/62/9.
4. Tsiramua, S. Logical-Probabilistic Modeling and Structural Analysis of Reconfigurable Systems Composed of Multifunctional Elements / S. Tsiramua, H. Meladze, T. Davitashvili // Pattern Recognition and Image Analysis. Advances in Mathematical Theory and Applications. – 2024. – Vol. 34, No. 1. – P. 144-157. – DOI 10.1134/S1054661824010218.
5. Rouban, A. I. The sensitivity coefficients for dynamic systems described by nonlinear difference generalized equations with the distributed memory and characterized by generalized functionals / A. I. Rouban // Tomsk State University Journal of Control and Computer Science. – 2022. – No. 61. – P. 81-87. – DOI 10.17223/19988605/61/8.

СВЕДЕНИЯ ОБ АВТОРАХ/ ABOUT THE AUTHORS

Татьяна Сергеевна Яницкая, кандидат технических наук, доцент, доцент «Высшей школы передовых производственных технологий», Поволжский государственный университет сервиса, 445017, Россия, г. Тольятти, ул. Гагарина, д. 4, yan@tolgas.ru
ORCID 0000-0001-7671-1763

Tatiana Sergeevna Ianitskaia, Candidate of Technical Sciences, Associate Professor, Professor at the Department of Higher School of Advanced Manufacturing Technologies, Volga Region State University of Service, Russia, Tolyatti, 4 Gagarin St., 445017

Руслан Ирекович Усманов, аспирант, Поволжский государственный университет сервиса, 445017, Россия, г. Тольятти, ул. Гагарина, д. 4, rslnusmnv@gmail.com

Ruslan Irekovich Usmanov, Post-Graduate Student, Volga Region State University of Service, Russia, Tolyatti, 4 Gagarin St., 445017