

Нургалеев Марат Фанисович

студент 2-го курса магистратуры, Федеральное государственное бюджетное образовательное учреждение высшего образования «Уфимский университет науки и технологий», Россия, Уфа

Бежаева Оксана Яковлевна

кандидат технических наук, доцент, Федеральное государственное бюджетное образовательное учреждение высшего образования «Уфимский университет науки и технологий», Россия, Уфа

ОБЪЯСНИМАЯ КАСКАДНАЯ СХЕМА ВОССТАНОВЛЕНИЯ ПРОПУСКОВ И ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ В СИСТЕМЕ ПРОИЗВОДСТВЕННОГО УЧЕТА НЕФТЕГАЗОДОБЫЧИ

Аннотация

В статье рассматривается задача объяснимого восстановления пропусков и поддержки принятия решений в корпоративной платформе производственного учета нефтегазодобычи, реализованной как микросервисная экосистема. Практическая значимость проблемы обусловлена неполными загрузками, поздними корректировками, расхождением агрегатов, повторной обработкой записей и контекстно-зависимыми аномалиями, которые не выявляются надежно с помощью простых правил валидации. В работе предлагается каскадная схема импутации, встроенная в более широкий контур контроля качества данных и объединяющая детерминированное восстановление, реконструкцию по временным рядам, восстановление по похожим объектам и рекомендательный режим для случаев с низкой уверенностью. Ключевой вклад статьи состоит в интеграции локальной объяснимости в операционный workflow: каждая проблемная запись преобразуется в структурированный DQ-кейс, содержащий тип нарушения, уровень уверенности, предложенное восстановленное значение, обоснование рекомендации и следующее рекомендуемое действие пользователя. Подход апробирован на обезличенной и частично синтетической выборке, моделирующей типовые противоречия производственного учета. Показано, что

детерминированные и временные стратегии остаются наиболее безопасными для критичных показателей, тогда как модельное восстановление особенно полезно при дефиците истории и сложных многомерных зависимостях. С практической точки зрения предложенное решение позволяет уменьшить объем ручной сверки, сократить время локализации источника расхождения и создать управляемый механизм использования методов машинного обучения в промышленной системе с высокой стоимостью ошибки.

Ключевые слова

качество данных; восстановление пропусков; объяснимый искусственный интеллект; производственный учет; нефтегазодобыча; микросервисная архитектура; выявление аномалий; поддержка принятия решений; временные ряды.

Введение

Крупные промышленные предприятия все в большей степени опираются на цифровые платформы производственного учета как на основной источник информации для оперативного планирования, анализа отклонений, мониторинга добычи и подготовки управленческой отчетности. В нефтегазодобыче такие платформы агрегируют факты, планы, ограничения, события, справочные атрибуты, прогнозные сценарии и отчетные показатели из нескольких операционных систем. Ценность подобной платформы определяется не только фактом наличия данных, но и их согласованностью, своевременностью и интерпретируемостью, поскольку именно эти характеристики влияют на перерасчет агрегатов, аналитические панели и управленческие решения. Если запись оказывается неполной либо семантически противоречивой, проблема редко остается локальной: она распространяется через границы сервисов и затрагивает последующие вычисления, сверки и пользовательские сценарии.

Настоящая статья посвящена именно задаче восстановления пропусков как части более общего контура интеллектуального контроля качества данных. Основная гипотеза состоит в том, что восстановление значений в промышленной

учетной системе не следует рассматривать как задачу единственной предиктивной модели. Напротив, оно должно быть организовано как управляемый каскад, в котором более прозрачные и менее рискованные стратегии всегда имеют приоритет над сложными статистическими и ML-методами. Такой подход особенно важен для микросервисных корпоративных систем, где модификации данных обязаны быть аудируемыми, бизнес-ответственность распределена между доменами, а полностью автоматическая корректировка критичных показателей может быть недопустимой.

Контекст задачи и обзор подходов

Несмотря на методическое разнообразие, прямой перенос общих техник *imputation* в промышленный производственный учет сталкивается с существенными ограничениями. Во-первых, стоимость ошибки здесь асимметрична: для одних индикаторов небольшое отклонение допустимо, тогда как для других неправильная автоматическая подстановка может исказить регламентированную либо управленческую отчетность. Во-вторых, пользователю требуется объяснение, почему предложенное значение считается разумным. В-третьих, сам процесс порождения данных содержит специфические артефакты: поздние импорты, повторно проигранные события, запаздывающие утверждения, конфликтующие версии перерасчета и рассогласование между пакетными и интерактивными обновлениями. Следовательно, механизм восстановления в промышленной системе должен быть контекстно-чувствительным, риск-ориентированным и тесно связанным с *workflow* подтверждения, отклонения и эскалации.

Отдельного внимания заслуживает направление *explainable artificial intelligence*. В прикладных индустриальных системах интерпретируемость является не дополнительной опцией, а обязательным условием внедрения. Предложенное системой значение становится операционно полезным только тогда, когда специалист понимает, получено ли оно из строгого предметного правила, из временного паттерна, из сравнения с похожей группой объектов или из вероятностной модели, обученной на истории исправлений. Поэтому в данной

работе объяснимость трактуется не как отдельный слой отчетности, добавленный после предсказания, а как внутренняя характеристика всего каскада восстановления.

Предлагаемая объяснимая каскадная схема восстановления

Предлагаемая схема включает четыре уровня восстановления, расположенные по мере возрастания аналитической сложности и операционного риска. Первый уровень представляет собой детерминированное восстановление. Если пропущенное либо сомнительное значение может быть выведено из балансового равенства, агрегатной зависимости, пересчета единиц измерения либо другого предметного правила с явной семантикой, такой вариант получает безусловный приоритет. Детерминированное восстановление обеспечивает максимальную степень доверия, поскольку пользователь может непосредственно проверить логику расчета, а система способна сохранить вместе с результатом точный след примененного правила.

Второй уровень связан с временным восстановлением. Он активируется в тех случаях, когда объект обладает достаточной и устойчивой историей, а дефект относится к короткому разрыву либо единичному выбросу. В качестве кандидатов используются значения, полученные с помощью регрессии по лагам, локальных медианных профилей, экспоненциального сглаживания и сопоставления с аналогичными периодами. Целью здесь является не долгосрочное прогнозирование, а восстановление правдоподобного краткосрочного значения, согласованного с недавним состоянием объекта. Временное восстановление сопровождается оценкой уверенности, которая зависит от стабильности ряда, величины остатка и наличия сопоставимых исторических ситуаций.

Третий уровень реализует восстановление по похожим объектам. Если индивидуальная история недостаточна либо нестабильна, система переходит к оценке на основе реер-group. Сходство определяется через предметные атрибуты: тип объекта, производственную зону, состояние эксплуатации, группу оборудования, сезонный контекст и характеристики плана. Табличная

supervised-модель, например градиентный бустинг, использует эти признаки вместе с процессными метками для построения кандидата восстановленного значения. Такой уровень особенно полезен для новых объектов, разреженных историй и тех ситуаций, где пропуск определяется комбинацией индивидуальных и поперечных факторов.

Четвертый уровень представляет собой рекомендательный режим. Если итоговая уверенность не достигает эксплуатационного порога, подсистема не изменяет учетную запись автоматически. Вместо этого она формирует DQ-кейс, содержащий один или несколько кандидатов, описание использованного метода, значения уверенности, ключевые объясняющие факторы и рекомендуемое дальнейшее действие. Подобное проектное решение принципиально важно для показателей с высокой бизнес-значимостью, поскольку сохраняет контроль пользователя и одновременно ускоряет аналитическую работу, необходимую для разрешения противоречия.

Экспериментальный сценарий и обсуждение результатов

Экспериментальная оценка проводилась на обезличенной и частично синтетической выборке, отражающей типовые дефекты производственного учета: пропуски суточных значений, повторные загрузки, поздние корректировки, несогласованные агрегаты и противоречивые состояния между связанными модулями. Сравнивались три стратегии: только детерминированное восстановление, детерминированное восстановление в сочетании с временным анализом и полный каскад, включающий восстановление по похожим объектам и рекомендательный режим. Анализ был ориентирован как на числовое качество реконструкции, так и на практическую полезность для реального процесса сверки. Числовое качество оценивалось по MAE и RMSE, тогда как операционная полезность анализировалась через ожидаемое снижение трудозатрат на ручную проверку и долю кейсов, разрешаемых без дополнительного сбора контекста.

Результаты сопоставления вариантов восстановления и ожидаемый эффект автоматизации сведены на рисунке 1.



Рисунок 1 - Метрики качества восстановления и оценка сокращения трудозатрат на ручную сверку

Результаты показывают, что детерминированное восстановление остается наиболее безопасным механизмом для структурно определенных показателей, тогда как временное восстановление существенно улучшает обработку коротких разрывов для стабильных объектов. Модельное восстановление по похожим объектам дает основной выигрыш в ситуациях, где локальной истории недостаточно либо пропуск зависит от многомерного контекста. В экспериментальном сценарии полный каскад снизил ожидаемую ошибку восстановления с 6,7 до 2,9 по MAE и с 11,2 до 4,8 по RMSE относительно простейшего baseline. Одновременно рекомендательный режим предотвращал неоправданную автоматизацию в случаях с низкой уверенностью. Такое сочетание повышения числового качества и сохранения управляемого участия пользователя критично для практического внедрения, поскольку успех индустриального AI определяется не только точностью, но и доверием, прозрачностью и соблюдением правил governance.

Заключение

Проведенное исследование показывает, что объяснимое восстановление пропусков для производственного учета нефтегазодобычи должно проектироваться как управляемый каскад, а не как единичный предиктивный алгоритм. Комбинация детерминированной логики, временной реконструкции, моделирования по похожим объектам и рекомендательного режима формирует сбалансированную схему, совместимую с микросервисной архитектурой промышленной системы и с эксплуатационной потребностью в аудируемости

решений. Предложенный подход охватывает как математическую сторону задачи восстановления, так и организационную сторону пользовательского принятия решений.

С инженерной точки зрения подход задает реалистичный путь к снижению трудоемкости ручной сверки, сохранению бизнес-контроля над чувствительными показателями и накоплению обратной связи для непрерывного улучшения моделей.

Список литературы

1. Little R. J. A., Rubin D. B. Statistical Analysis with Missing Data. 3rd ed. Hoboken: John Wiley and Sons, 2019. 462 p.
2. Hyndman R. J., Athanasopoulos G. Forecasting: Principles and Practice. 3rd ed. Melbourne: OTexts, 2021. 442 p.
3. Chandola V., Banerjee A., Kumar V. Anomaly Detection: A Survey // ACM Computing Surveys. 2009. Vol. 41. No. 3. Article 15.
4. Kuhn M., Johnson K. Applied Predictive Modeling. New York: Springer, 2013. 600 p.
5. Ribeiro M. T., Singh S., Guestrin C. Why Should I Trust You?: Explaining the Predictions of Any Classifier // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016. P. 1135-1144.
6. Lundberg S. M., Lee S.-I. A Unified Approach to Interpreting Model Predictions // Advances in Neural Information Processing Systems. 2017. Vol. 30. P. 4765-4774.
7. Kimball R., Ross M. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. 3rd ed. Indianapolis: Wiley, 2013. 600 p.
8. Provost F., Fawcett T. Data Science for Business. Sebastopol: O'Reilly Media, 2013. 414 p.