

УДК 004

***МАСШТАБИРУЕМЫЕ МЕТОДЫ ОБУЧЕНИЯ С ЧАСТИЧНО
РАЗМЕЧЕННЫМИ КОРПОРАТИВНЫМИ ДАННЫМИ В УСЛОВИЯХ
НЕОДНОРОДНЫХ БИЗНЕС-ПРОЦЕССОВ***

Тохириён Б.,

кандидат технических наук, доцент,

ФГБОУ ВО «Уральский государственный экономический университет»

г. Екатеринбург, Российская Федерация.

Аннотация

Статья посвящена анализу масштабируемых методов обучения с частично размеченными корпоративными данными в условиях неоднородных бизнес-процессов. Рассматриваются теоретические основы полусупервизорного обучения, влияние структурной и динамической гетерогенности бизнес-процессов на устойчивость и переносимость моделей, а также архитектурные и методологические решения, обеспечивающие масштабируемость и управляемость таких моделей в корпоративной среде. Обоснована необходимость интеграции алгоритмов с системами процессной аналитики, онтологиями данных и корпоративными механизмами управления качеством и комплаенсом.

Ключевые слова: частичная разметка, корпоративные данные, полусупервизорное обучение, масштабируемость, бизнес-процессы.

***SCALABLE LEARNING METHODS WITH PARTIALLY LABELED
CORPORATE DATA IN HETEROGENEOUS BUSINESS PROCESSES***

Tokhiriyon B.,

PhD (Engineering), Associate Professor,

Ural State University of Economics

Yekaterinburg, Russian Federation.

Abstract

This article analyzes scalable learning methods with partially labeled corporate data in heterogeneous business processes. It examines the theoretical foundations of semi-supervised learning, the impact of structural and dynamic heterogeneity of business processes on the stability and portability of models, and architectural and methodological solutions that ensure the scalability and manageability of such models in a corporate environment. The need for integrating algorithms with process analytics systems, data ontologies, and corporate quality and compliance management mechanisms is substantiated.

Keywords: partial tagging, corporate data, semi-supervised learning, scalability, business processes.

Цифровая трансформация крупных предприятий приводит к лавинообразному росту объемов корпоративных данных, распределенных между разнородными информационными системами, хранилищами и прикладными сервисами, что порождает фрагментированную корпоративную архитектуру данных и высокую вариативность сквозных и межорганизационных бизнес-процессов. При этом трудоемкость и стоимость экспертизы делают полноту ручной разметки практически недостижимой, поэтому в реальных корпоративных контурах доминируют частично размеченные выборки, дополняемые массивами неразмеченных событийных логов и документов. Теоретические основы обучения с частично размеченными данными были системно заложены Оливье Шапелем, Бернхардом Шёлькопфом и Александром Циеном, показавшими, что привлечение неразмеченных наблюдений при соблюдении определенных структурных допущений позволяет повышать обобщающую способность моделей и предложившими таксономию соответствующих алгоритмических классов. В отечественной научной школе работы Александра Дьяконова подчеркивают роль геометрии пространства

признаков и распределения неразмеченных объектов в задачах самообучения и обучения с частичной разметкой, что особенно актуально для высокоразмерных корпоративных признаков пространств.

Одновременно исследования в области процессно-ориентированного управления и процессной аналитики, развиваемые, в частности, Вилом ван дер Аалстом, демонстрируют, что гетерогенные бизнес-процессы могут быть формализованы на основе журналов событий и варьирующихся конфигураций рабочих потоков, что задает дополнительные ограничения к переносимости и масштабируемости моделей. Российские коллективы, такие как группа Алексея Якушева и соавторов, уже демонстрируют применимость подходов обучения на частично размеченных данных к корпоративным текстовым коллекциям и задачам извлечения сущностей в условиях ограниченных ресурсов разметки [6, с. 5].

В теории обучения с частичной разметкой рассматривается смешанная информационная структура, в которой ограниченный эталонный корпус размеченных объектов дополняется крупными массивами неразмеченных наблюдений; при этом целевая функция качества выводится с опорой на априорные гипотезы гладкости решающей функции, кластерного предположения, гипотезы многообразия и разделения классов по областям низкой плотности. Концептуальная конструкция таких подходов системно изложена в работах Оливье Шапеля, Бернхарда Шёлькопфа и Александра Циена, предложивших таксономию генеративных и дискриминативных схем, графовых методов на основе лапласианов смежности и алгоритмов с распределительно-зависимой регуляризацией.

В рамках данной парадигмы корпоративная среда трактуется как особый класс источников данных с высокой степенью институциональной и процессной неоднородности: журналы событий бизнес-процессов, транзакционные потоки, документооборот и коммуникационные ленты формируют многоуровневое пространство признаков, подчинённое регламентам управления качеством

Дневник науки | www.dnevniknauki.ru | СМИ Эл № ФС 77-68405 ISSN 2541-8327

данных, архитектурным ограничениям корпоративных хранилищ и витрин, а также требованиям стратегического управления на основе данных. В таких условиях частичная разметка возникает как следствие ограниченности экспертизы и разнородности локальных регламентов в подразделениях, что стимулирует использование самообучения, кообучения, методов работы со слаборазмеченными данными и гибридных схем, совмещающих вероятностные модели и функционально-аналитическую регуляризацию, позволяющих переносить знания между сегментами корпоративной экосистемы при сохранении согласованности с онтологией предметной области и политиками управления данными [2, с. 83].

Неоднородность бизнес-процессов в крупных корпоративных экосистемах проявляется в многообразии вариантов маршрутов выполнения, регламентов, организационно-ролевых конфигураций и информационных каналов, что приводит к формированию смешанных и нестационарных распределений данных, нарушающих стандартные предпосылки независимости и однородности выборки для моделей машинного обучения. На уровне журналов событий это выражается в сосуществовании семейств процессных вариантов с различной глубиной ветвления, асимметрией циклов и неодинаковой насыщенностью контекстными атрибутами, в результате чего единая модель фактически обучается на совокупности перекрывающихся подсистем с разными латентными механизмами генерации траекторий. Исследования в области интеллектуального анализа бизнес-процессов под руководством Вила ван дер Аалста и его коллег показывают, что изменение структур потоков работ, ресурсов и данных приводит к дрейфу концепции, когда ранее построенные модели теряют прогностическую и диагностическую пригодность даже при сохранении внешне схожих показателей производительности.

«В предиктивном мониторинге бизнес-процессов неоднородность проявляется в чувствительности моделей к скрытой сегментации клиентов, каналов обслуживания и территориальных подразделений: одни и те же признаки

Дневник науки | www.dnevniknauki.ru | СМН ЭЛ № ФС 77-68405 ISSN 2541-8327

могут иметь противоположную прогностическую роль в различных сегментах, что приводит к систематическим смещениям и деградации точности при переносе моделей между контурами. Работы, посвящённые прогнозированию последующих действий и временных характеристик для нестационарных процессов, демонстрируют, что появление ранее не наблюдавшихся комбинаций атрибутов и новых локальных регламентов существенно снижает устойчивость моделей и требует специализированных стратегий кодирования признаков, адаптивного дообучения и периодической переидентификации процессных паттернов» [4, с. 33].

Современные подходы к обнаружению и реагированию на дрейф концепции в бизнес-процессах трактуют его как задачу последовательного управленческого выбора между пересмотром модели, изменением политики отбора данных и корректировкой самих регламентов, что подчёркивает взаимосвязанность алгоритмической и организационной архитектур. С точки зрения корпоративного управления неоднородность бизнес-процессов обусловлена диверсификацией продуктового портфеля, матричной структурой подотчётности, различиями в зрелости локальных контуров управления и неодинаковыми практиками ведения справочников и журналов событий, вследствие чего совокупный массив данных приобретает выраженную иерархическую, многоисточниковую и контекстно-зависимую структуру.

В таких условиях модели машинного обучения, игнорирующие процессную и организационную стратификацию, приводят к управленческим решениям с асимметричным рисковым профилем: часть подразделений получает заведомо некорректные прогнозы и рекомендации, ухудшается интерпретируемость результатов для владельцев процессов, возрастает вероятность конфликта между показателями эффективности и фактической операционной реальностью. «Поэтому неоднородность бизнес-процессов должна рассматриваться как фундаментальное ограничение при проектировании масштабируемых алгоритмов: необходимы встроенные механизмы сегментации

Дневник науки | www.dnevniknauki.ru | СМИ Эл № ФС 77-68405 ISSN 2541-8327

иерархий процессов, учёт контекстных метаданных об организационной структуре и регламентах, а также тесная интеграция контуров аналитического моделирования с контурами управления изменениями и процессной архитектурой предприятия» [5, с. 184].

Масштабируемые архитектуры обучения с частично размеченными корпоративными данными в условиях неоднородных бизнес-процессов опираются на многослойную организацию вычислительной и управленческой инфраструктуры. На нижнем уровне формируется распределённая архитектура корпоративных хранилищ и озёр данных с поддержкой параллельной загрузки событийных журналов, транзакций и документопотоков, а также их инкрементальной очистки и нормализации. Над ним располагается семантический слой с онтологиями предметных областей, справочниками мастер-данных и моделями бизнес-процессов, обеспечивающий согласованность признаков пространств при объединении неоднородных источников. Верхний уровень образуют контуры построения, валидации и промышленной эксплуатации моделей, в которых реализуются стандартизованные конвейеры подготовки выборок, обучения, мониторинга дрейфа концепции и управляемого переразвертывания. Современные обзоры распределённого и федеративного обучения подчёркивают, что для масштабируемой работы с частично размеченными выборками критично использование схем параллелизма по данным и по параметрам, а также федеративных протоколов, позволяющих агрегировать знания между дочерними обществами и странами без физической консолидации персональных данных.

В методологическом плане особое место занимает класс графовых и многошаговых алгоритмов полусупервизорного обучения, предназначенных для работы с крупномасштабными корпоративными графами — сетями клиентов, контрагентов, продуктов, технологических операций и событийных связей. Специализированные обзоры подчёркивают, что повышение вычислительной эффективности таких методов достигается за счёт приближённых спектральных

Дневник науки | www.dnevniknauki.ru | СМИ Эл № ФС 77-68405 ISSN 2541-8327

разложений, иерархического уплотнения графов, стохастической выборки подграфов и локализованных вариантов регуляризации по гладкости на многообразиях, что делает возможной обработку миллионов вершин и рёбер в рамках единого контура обучения с частичной разметкой. В корпоративной среде это позволяет строить модели распространения риска, выявления аномальных цепочек событий и сегментации клиентов с опорой на ограниченный набор размеченных узлов, распространяющий метки по структуре бизнес-взаимодействий, а не только по индивидуальным признакам объектов.

Важным направлением являются архитектуры слабого надзора, в которых вместо трудоёмкой ручной разметки используются программируемые источники слабо формализованных правил, эвристик и внешних знаний. Платформенные решения класса «Сноркел» и их промышленное развитие демонстрируют, что путём агрегирования множества признаков и логических функций-разметчиков, моделирования их надёжности и разрешения противоречий можно автоматически генерировать крупные псевдоразмеченные корпуса, сопоставимые по полезности с десятками тысяч вручную размеченных примеров, и при этом масштабировать процесс разметки до миллионов экземпляров. «Для корпоративных данных это открывает возможность встроить этап конструирования правил слабого надзора в управленческие процедуры: бизнес-эксперты, владельцы процессов и специалисты по комплаенсу формализуют свои регламенты и эвристики в виде библиотек функций-разметчиков, которые затем централизованно калибруются и используются в конвейерах обучения» [7, с. 3].

Отдельный класс масштабируемых решений базируется на микросервисных архитектурах аналитических платформ, где задачи подготовки данных, обучения моделей, актуализации параметров и онлайн-применения разделены на автономные сервисы с чёткими интерфейсами. Исследования по распределённой классификации в режиме полусупервизорного обучения показывают, что подобная декомпозиция позволяет эластично масштабировать

Дневник науки | www.dnevniknauki.ru | СМИ ЭЛ № ФС 77-68405 ISSN 2541-8327

отдельные стадии конвейера, использовать адаптивное перераспределение вычислительной нагрузки и обновлять модели без остановки основного производственного контура. В корпоративной практике такие архитектуры сочетаются с оркестраторами рабочих потоков, системами управления версиями моделей и политиками согласованного развёртывания, что обеспечивает управляемость и прослеживаемость изменений при работе с неоднородными бизнес-процессами.

С точки зрения методологии обучения с частичной разметкой ключевую роль играет интеграция нескольких комплементарных механизмов: самообучения с псевдоразметкой, активного обучения с приоритизацией объектов для экспертной верификации, слабого надзора на основе корпоративных регламентов и графовых методов распространения меток. В предиктивном мониторинге бизнес-процессов это реализуется через специализированные архитектуры глубинного обучения над журналами событий, которые сочетают рекуррентные и сверточные компоненты, механизмы внимания и временные кодировщики, позволяя масштабно предсказывать дальнейшее поведение кейсов, сроки завершения и риск нарушений соглашений об уровне сервиса даже при неполной разметке исторических данных. Такие архитектуры дополняются процедурами стратификации по типам процессов и организационным подразделениям, что позволяет учитывать неоднородность контуров и избегать систематического смещения прогнозов.

Таким образом, обучение с частичной разметкой в корпоративной среде целесообразно рассматривать как элемент целостной архитектуры управления данными и бизнес-процессами. Неоднородность процессов, дрейф концепции и фрагментация источников информации требуют масштабируемых алгоритмов, сочетающих графовые методы, слабый надзор, самообучение и активную экспертизу. Обоснована необходимость интеграции таких алгоритмов с процессной аналитикой, онтологиями предметных областей и корпоративными контурами комплаенса. Перспективными направлениями дальнейших

Дневник науки | www.dnevniknauki.ru | СМН ЭЛ № ФС 77-68405 ISSN 2541-8327

исследований являются формализация критериев стратификации процессов, построение мультиуровневых моделей переноса знаний между подразделениями и разработка управленческих регламентов жизненного цикла аналитических моделей. Это позволит связать технические решения с практиками стратегического контроллинга, риск-менеджмента и проектирования целевых операционных моделей в условиях высокой турбулентности среды.

Библиографический список:

1. Давыдов С.Г., Матвеева Н.Н., Адемукова Н.В., Вичканова А.А. Искусственный интеллект в российском высшем образовании: текущее состояние и перспективы развития / С.Г. Давыдов [и др.] // Университетское управление: практика и анализ. – 2024. – Т. 28. – № 3. – С. 32-44.
2. Карачун И.А. Машинное обучение в сфере платформизации бизнес-моделей / И.А. Карачун // Журнал Белорусского государственного университета. Экономика. – 2022. – № 1. – С. 79-88.
3. Кугаевских А.В., Муромцев Д.И., Кирсанова О.В. Классические методы машинного обучения. – СПб.: Университет ИТМО, 2022. – 53 с.
4. Лебедев И.С. Обработка информационных последовательностей с использованием адаптивного анализа сегментов при оценке состояния систем / И.С. Лебедев // Информационно-управляющие системы. – 2025. – № 3. – С. 25-36.
5. Панина О.В., Попадюк Н.К., Еремин С.Г., Токмурзин Т.М., Разумова Е.В. Применение технологий BigData для оптимизации производственных процессов в горнодобывающей промышленности России: анализ внедрения и эффективности // Горная промышленность. – 2024. – № 6. – С. 178-185.
6. Савин С.В., Мурзин А.Д. Искусственный интеллект как инструмент управления инновационными проектами и стартапами // Матер. XIV Всерос. совещания по проблемам управления «ВСПУ-2024» (Москва, 17–20 июня 2024 г.). – Москва, 2024. – С. 1-6.

7. Суханов К.А., Жижома А.И. Адаптивное обучение математике и IT с использованием технологий машинного обучения // Матер. XVI Всерос. науч.-практ. конф. молодых ученых «Россия молодая» (Ставрополь, 16–19 апреля 2024 г.). – Ставрополь, 2024. – С. 1-5.
8. Харьков В.П. Информационные технологии в бизнесе: практический курс для менеджеров и экономистов. – Москва: Изд-во Московского гуманитарного университета, 2025. – 98 с.
9. Хасан Р.А., Мурзина Э.Ф., Ризванова М.А. Технологии искусственного интеллекта как бустер развития малого бизнеса // Экономика и управление: научно-практический журнал. – 2024. – № 3. – С. 54-60.

Оригинальность 80%