Кластерно-гибридная модель товарных рекомендаций для корпоративных онлайн-закупок

Моисеев Никита Александрович
доктор экономических наук, профессор кафедры математических методов в экономике, Российский экономический университет им. Г.В. Плеханова, moiseev.na@rea.ru,

Одоевский Иван Игоревич
аспирант кафедры математических методов в экономике, Российский экономический университет им. Г.В. Плеханова, odoivan98@gmail.com

Объектом исследования выступают рекомендательные системы корпоративных электронных торговых платформ, предметом – методы ранжирования товарных категорий при разреженной истории закупок и неоднородности клиентской базы. Цель работы состоит в разработке и вычислительной проверке кластерно-гибридной модели, объединяющей item-based коллаборативную фильтрацию, групповой профиль K-Prototypes, отраслевой приор и штраф за глобальную популярность. Эксперимент проведён на полностью воспроизводимой синтетической панели из 2000 корпоративных клиентов и 25 000 транзакций по 12 товарным категориям. События каждого клиента разделены во времени на обучающую, валидационную и тестовую части в пропорции 60/20/20. Качество оценивалось по Precision@5, Recall@5, NDCG@5, охвату каталога, доле непопулярных категорий и нормированной энтропии распределения экспозиций. Предложенная модель достигла NDCG@5 = 0,4650 против 0,4068 у ранжирования по популярности, 0,4029 у кластерного метода и 0,3464 у item-based коллаборативной фильтрации. Прирост относительно сильнейшего базового решения составил 0,0582; 95%-й бутстреп-интервал [0,0470; 0,0695]. При этом нормированная энтропия экспозиций увеличилась с 0,6477 до 0,9067, а доля категорий «длинного хвоста» – с нуля до 0,2447. Научная новизна состоит в сочетании сегментации, персонализации и коррекции смещения популярности в одном ранжирующем контуре. Результаты подтверждают работоспособность метода в контролируемой среде, но не являются оценкой производственного эффекта без апробации на реальных данных.

Ключевые слова: B2B-электронная коммерция, K-Prototypes, неявная обратная связь, смещение популярности, энтропия экспозиции, синтетические данные, ранжирование.

Финансирование. Данное исследование выполнено в рамках государственного задания в сфере научной деятельности Министерства науки и высшего образования РФ на тему «Методы и алгоритмы искусственного интеллекта в решении задач моделирования и прогнозирования динамики сложных социально-экономических систем», номер проекта FSSW-2026-0004.

Введение

Развитие цифровых торговых платформ постепенно изменяет саму логику межфирменной торговли. Если первые B2B-витрины фактически переносили в сеть бумажный каталог, то современная платформа должна сокращать издержки поиска, учитывать профиль организации и предлагать следующее целесообразное действие.

Несмотря на внешнее сходство интерфейсов электронной коммерции, закупки в сегменте B2B подчинены другой институциональной логике. Вектор взаимодействия здесь смещается в сторону многоуровневого утверждения сделок, параметры которых строго детерминированы генеральными соглашениями, специальными тарифными сетками и графиками снабжения. Различия между B2C и B2B подробно рассматриваются в отечественных работах по электронной торговле [6].

Исторически большинство алгоритмов развивалось на данных B2C. Матричная факторизация, коллаборативная фильтрация и парное ранжирование стали фундаментом для фронтирных систем [7; 15; 17; 5], а гибридизация закрепилась как способ соединить разнородные источники информации [8].

Первые специализированные работы по B2B предлагали строить систему вокруг бизнес-правил, доменных знаний и интеграции с корпоративными процессами [21], что привело к появлению гибридных моделей доверия, объединяющие сигналы от пользователей и товаров [19], а также графовых архитектур для поиска потенциальных клиентов [10]. Так, проект IndiaMART важен тем, что показывает не только алгоритм, но и эксперимент: доля принятых запросов в первых 25 позициях там выросла на 8,2% [12]. Этот результат нельзя однозначно и механически переносить на российские каталоги, но он задаёт правильное логику доказательства: офлайн-метрики должны завершаться контролируемой проверкой.

В современных разработках российских ученых последних лет акцент также смещается от общих рассуждений к количественно-измеримым моделям [2; 3; 4]. Однако, методический разрыв между сегментацией клиентов, персональным ранжированием и управлением смещением популярности остаётся недостаточно формализованным.

Отдельной проблемой является смещение в сторону бестселлеров. На данных с неравномерным спросом модель склонна снова и снова возвращать наиболее популярные позиции в списке номенклатуры. В итоге алгоритм усиливает историческую концентрацию [16]. Для B2B-каталога это означает, что нишевая запасная часть или отраслевой расходник могут быть исключены из видимой части рекомендаций, хотя именно они генерируют выручку для узкого сегмента. Поэтому оценка должна принимать во внимание не только точность, но и разнообразие, новизну и охват [14].

Энтропия Шеннона – функционал от вектора вероятностей, который характеризует неопределённость, а не заменяет само распределение [20]. В настоящей работе она используется строго как системная метрика равномерности распределения экспозиций. Она не трактуется как автоматический индикатор полезности отдельной рекомендации.

Так, кластерный профиль, индивидуальная история и отраслевой контекст обычно описываются раздельно, хотя каждый из этих сигналов отвечает на разный вопрос: что покупают схожие клиенты, что покупал конкретный клиент и что в принципе характерно для его отрасли. Гипотеза авторов состоит в том, что их линейная комбинация с явной коррекцией популярности повышает точность следующей покупки без резкого снижения разнообразия выдачи.

Цель данного исследования – разработать и проверить кластерно-гибридную модель товарных рекомендаций для B2B-каталога. Объектом исследования являются процессы персонализированного ранжирования товарных категорий на корпоративной платформе, а предметом – метод объединения поведенческого, кластерного и отраслевого скоров. Для достижения цели формализуются данные и алгоритм, выполняется временное тестирование, сравниваются четыре метода, и оценивается устойчивость результата при пяти независимых случайных зёрнах.

Научная новизна работы заключается в трёх взаимосвязанных элементах. Во-первых, кластеризация будет рассматриваться не как однозначная замена коллаборативной фильтрации, а как структурный приор. Во-вторых, в одном скоре будет учитываться история совместных покупок, профиль кластера и отраслевая структура спроса. В-третьих, коррекция популярности оценивается одновременно как по ранговым метрикам, так и по распределению экспозиций. Практическая значимость состоит в том, что модель может быть реализована как прозрачный слой повторного ранжирования поверх уже существующего кандидатного механизма.

Материалы и методы

Дизайн исследования и данные

Исследование построено как эксперимент с синтетической панелью транзакций. При таком выборе отсутствует возможность апробации на данных компании, но позволяет, и однозначно описать процедуру, и сохранить контроль над структурой данных, и проверить саму алгоритмическую гипотезу.

Панель содержит 2000 клиентов и 25 000 последовательных транзакций по 12 товарным категориям, для каждой организации заданы отрасль, федеральный округ, размер по выручке, уровень цифровой зрелости и индивидуальная склонность к возвратам, отраслевые и региональные доли используются как сценарные параметры, а не как эмпирическая оценка генеральной совокупности. Отчёт Data Insight привлекается для обоснования масштаба и неоднородности рынка, но не рассматривается как источник микроданных [1].

Транзакционная активность зависит от размера компании и цифровой зрелости. Каждому клиенту гарантировано не менее семи событий; оставшиеся события распределены пропорционально индексу активности. Вероятность выбора товарной категории складывается из четырёх компонентов: латентного архетипа закупок с весом 0,48, отраслевого профиля с весом 0,32, индивидуального дирихлевского вектора с весом 0,15 и глобальной популярности с весом 0,05. Данная конструкция создаёт и групповую закономерность, и индивидуальное отклонение от неё.

Для каждого клиента события упорядочены по месяцу и внутреннему номеру. Первые 60% используются для обучения, следующие 20% – для выбора весов гибридной модели, последние 20% – для единственной итоговой оценки. Случайное перемешивание строк не используется, так как оно привело бы к утечке информации о будущих покупках. После настройки обучающая и валидационная части объединяются, модель переобучается, а тестовые события остаются неприкосновенными.

Таблица 1
Параметры синтетической панели

Параметр	Значение	Назначение
Клиенты / транзакции	2 000 / 25 000	Контролируемый масштаб и разреженность
Товарные категории	12	Объекты ранжирования
Отрасли / регионы / размеры	8 / 8 / 4	Категориальный профиль организации
Числовые признаки	5	Частота, чек, объём, зрелость, возвраты
Временное разделение	60/20/20	Обучение, валидация, отложенный тест
Случайное зерно	20260813	Воспроизводимость основного прогона

Источник: разработано авторами.

Кластеризация клиентов

Клиентский профиль содержит пять числовых признаков: логарифм числа транзакций, логарифм среднего чека, логарифм общего объёма закупок, цифровую зрелость и наблюдаемую долю возвратов. Перед K-Means признаки стандартизуются. Алгоритм K-Prototypes дополнительно включает отрасль, регион и размер компании. Он минимизирует сумму квадратов евклидовых расстояний по числовой части и число несовпадений по категориальной. Коэффициент несовпадения задан равным единице. Такая метрика следует исходной логике K-Prototypes [13].

$$d(u,c) = \Sigma_j (z_{uj} - q^c_j)^2 + \eta \Sigma_l 1(x_{ul} \neq q^c_l), \quad \eta = 1 \quad (1)$$

В формуле (1) z_{uj} – стандартизованное значение j-го числового признака клиента u, q^c_j – соответствующая координата прототипа кластера c, x_{ul} и q^c_l – категориальные значения. При пересчёте прототипа для числовых признаков берётся среднее, для категориальных – мода. Для каждого k от 2 до 8 выполняется не менее шести случайных инициализаций; сохраняется решение с минимальной целевой функцией. Число кластеров выбирается по максимуму среднего коэффициента силуэта [18], а индекс Дэвиса–Болдина используется как дополнительная мера компактности и разделимости [11].

Алгоритмы ранжирования

Для корректного сравнения сформированы три базовых метода. Первый возвращает глобально популярные категории и является трудным ориентиром для рынка с высокой долей повторных закупок. Второй представляет собой item-based коллаборативную фильтрацию: косинусное сходство между категориями рассчитывается по бинарной матрице «клиент–категория», после чего профиль клиента умножается на матрицу сходств. Третий метод ранжирует категории по частоте внутри кластера K-Prototypes. Данные методы позволяют отделить вклад популярности, индивидуального поведения и группового профиля.

Авторская модель дополняет эти сигналы отраслевым приором и штрафом за глобальную популярность. Каждый исходный скор нормируется внутри клиента на отрезке [0; 1]. Популярность преобразуется логарифмом, чтобы одна крупная категория не доминировала над остальными. Такой штраф не объявляет нишевую позицию релевантной автоматически; он только снижает избыточное преимущество частотного сигнала.

$$s_{ui} = \alpha \cdot CF_{ui} + \beta \cdot CL_{ui} + (1-\alpha-\beta) \cdot IND_{ui} - \lambda \cdot POP_i \quad (2)$$

В формуле (2) CF_{ui} – нормированный item-based скор, CL_{ui} – частота категории i в кластере клиента u, IND_{ui} – отраслевой приор, POP_i – нормированный логарифм глобальной частоты. Параметры перебираются по сетке $\alpha \in \{0,35; 0,50; 0,65\}$, $\beta \in \{0,15; 0,30; 0,45\}$, $\lambda \in \{0; 0,04; 0,08; 0,12\}$ при $\alpha + \beta < 0,95$. Критерий валидации равен NDCG@5 + 0,03·Hnorm. В основном прогоне выбраны $\alpha = 0,35$, $\beta = 0,15$ и $\lambda = 0,08$; следовательно, вес отраслевого приора равен 0,50.

Метрики и статистическая проверка

Релевантной считается товарная категория, встретившаяся в тестовых транзакциях клиента. Поскольку в B2B повторная закупка является нормальным сценарием, ранее встречавшиеся категории не исключаются из кандидатов. Precision@5 показывает долю релевантных категорий в первых пяти позициях, Recall@5 – долю охваченных тестовых категорий, NDCG@5 – качество их порядка с большим весом верхних позиций. Использование нескольких метрик снижает риск подмены качества ранжирования одной удобной цифрой [14].

$$Precision@5 = |R_u^5 \cap T_u| / 5; \quad Recall@5 = |R_u^5 \cap T_u| / |T_u| \quad (3)$$

$$NDCG@5 = DCG@5 / IDCG@5, \quad DCG@5 = \Sigma_{r=1}^{5} rel_r / \log_2(r+1) \quad (4)$$

Разнообразие оценивается на уровне всей выдачи. Пусть p_i – доля i-й категории среди всех позиций Top-5. Тогда нормированная энтропия равна энтропии Шеннона, делённой на логарифм числа категорий. Значение, близкое к единице, означает более равномерное распределение экспозиций, но не гарантирует релевантность. Дополнительно рассчитываются охват каталога и доля «длинного хвоста». В эксперименте к нему операционально отнесена половина категорий с наименьшей частотой в обучающей части.

$$Hnorm = -\Sigma_i p_i \ln(p_i) / \ln(|I|) \quad (5)$$

Неопределённость средних ранговых метрик оценивается непараметрическим бутстрепом по клиентам. Для каждого сравнения выполняется 3000 повторных выборок с возвращением; приводится 95%-й процентильный интервал. Кроме того, весь цикл генерации и оценки повторяется для пяти зёрен от 20260813 до 20260817. Эта процедура проверяет, что результат не является случайной особенностью одной синтетической реализации.

Результаты исследования

Структура клиентских профилей

По коэффициенту силуэта наилучшим оказалось разбиение на 2 кластера. Для K-Means коэффициент силуэта равен 0,3416, индекс Дэвиса–Болдина – 1,0944; для K-Prototypes – 0,2505 и 1,4408 соответственно.

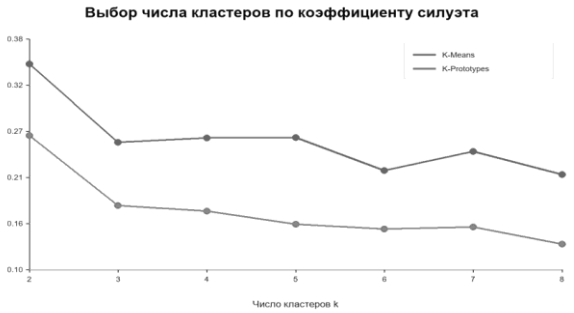


Рисунок 1 – Выбор числа кластеров по коэффициенту силуэта
Источник: разработано авторами на основе вычислительного эксперимента.

Полученные значения не дают оснований говорить о сильно изолированных сегментах: категориальные признаки увеличивают внутрикластерную неоднородность. Тем не менее разбиение полезно как групповой приор и не претендует на естественную типологию всего B2B-рынка.

Таблица 2
Профили кластеров K-Prototypes

Кластер	n	Транзакции	Средний чек, руб.	Цифровая зрелость	Размер	Отрасль
1	851	11,18	576 970	3,71	средняя	Машиностроение
2	1149	7,76	177 972	2,43	малая	Машиностроение

Источник: разработано авторами.

Качество ранжирования

Таблица 3
Метрики рекомендательных методов на тесте

Метод	P@5	R@5	NDCG@5	Охват	Hnorm	Long-tail
Популярность	0,2780	0,4887	0,4068	0,4167	0,6477	0,0000
Item-based CF	0,2397	0,4135	0,3464	0,9167	0,9415	0,3445
Кластер	0,2753	0,4820	0,4029	0,5000	0,7026	0,1149
Гибрид	0,3135	0,5514	0,4650	1,0000	0,9067	0,2447

Источник: разработано авторами.

Гибридная модель показала наилучший результат по всем трём ранговым метрикам: Precision@5 = 0,3135, Recall@5 = 0,5514, NDCG@5 = 0,4650. Относительно ранжирования по популярности NDCG@5 увеличилась на

14,3%, относительно кластерного метода – на 15,4%, относительно item-based CF – на 34,2%.

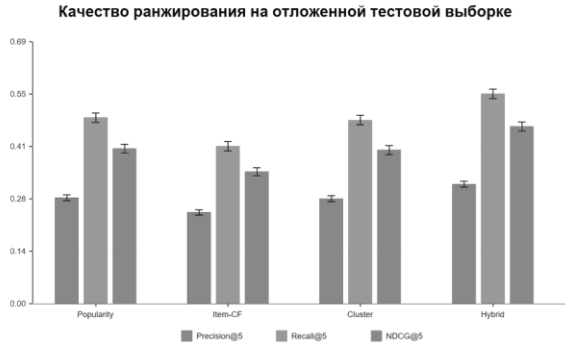


Рисунок 2 – Качество ранжирования на отложенной тестовой выборке
Источник: разработано авторами

Сам по себе кластерный метод не превзошёл популярность. Таким образом, гипотеза подтверждается не для «кластеризации вообще», а для её конкретной роли в гибридном скоре.

Таблица 4
Бутстреп-сравнение NDCG@5

Сравнение	Разность	95%-й интервал	pboot
Гибрид – Item-CF	0,1186	[0,1069; 0,1303]	< 0,001
Гибрид – популярность	0,0582	[0,0470; 0,0695]	< 0,001
Гибрид – кластер	0,0620	[0,0513; 0,0729]	< 0,001

Источник: разработано авторами.

Для всех трёх сравнений 95%-е интервалы не включают нуль. Наименьшая разность получена относительно сильного базового ранжирования по популярности: 0,0582 при интервале [0,0470; 0,0695]. Данный результат свидетельствует о внутренней устойчивости эффекта в пределах заданной генерирующей модели, но не позволяет автоматически переносить размер эффекта на реальную платформу.

Разнообразие экспозиций и устойчивость

Охват гибридной модели по каталогу равен 100,0%, нормированная энтропия – 0,9067, доля категорий «длинного хвоста» – 0,2447. Ранжирование по популярности охватывает только 41,7% каталога, имеет энтропию 0,6477 и не предоставляет ни одной категории из заданного «длинного хвоста». Item-based CF показал неожиданно даже более равномерную экспозицию (0,9415) и большую долю непопулярных категорий (0,3445), но существенно уступает по NDCG@5. Из этого следует, что ценность гибрида состоит не в максимизации одной метрики, а в балансе точности и охвата.

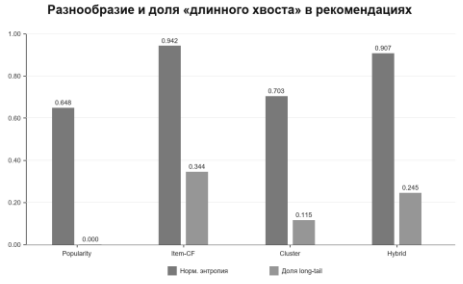


Рисунок 3 – Разнообразие экспозиций и доля «длинного хвоста»
Источник: разработано авторами на основе вычислительного эксперимента.

Таблица 5
Устойчивость гибридной модели по пяти независимым генерациям

Показатель	P@5	R@5	NDCG@5	Hnorm	Long-tail
Гибрид, среднее	0,3159	0,5547	0,4657	0,9051	0,2380
Стандартное отклонение	0,0020	0,0030	0,0033	0,0040	0,0053

Источник: разработано авторами.

Низкая вариация между семплами данных демонстрирует, что вывод не зависит от одного выбранного случайного зерна на удачу.

Обсуждение результатов

Первый важный вывод касается того, как следует понимать кластеризацию. Сам по себе кластер ещё не является рекомендацией. Он лишь показывает, к какой группе похожих клиентов относится организация. Это полезно, когда по конкретному клиенту мало данных: вместо нескольких его покупок можно использовать информацию о поведении всей группы. Но здесь возникает очевидный риск: если клиент заметно отличается от большинства участников своего кластера, рекомендации, основанные на поведении группы, могут оказаться для него нерелевантными.

Описание кластера через его прототип помогает понять, какой тип клиента в нём преобладает, но не делает всех клиентов внутри группы одинаковыми [9]. Из-за чего в итоговой модели кластерная частота используется только как один из сигналов и имеет вес 0,15. Если бы рекомендации строились лишь на ней, модель слишком часто подменяла бы индивидуальное поведение клиента поведением его группы.

Второй вывод состоит в том, что разные компоненты модели должны отвечать на разные вопросы. Item-based CF показывает, какие товарные категории часто встречаются вместе в покупках. Кластерный приор отвечает на другой вопрос: что обычно покупают организации, похожие по масштабу и уровню цифровой зрелости. Отраслевой приор добавляет ещё один слой информации — какие товары логичны именно для данного сектора экономики. Поэтому преимущество гибридной модели возникает не потому, что она просто сложнее. Эффект появляется тогда, когда каждый компонент использует собственный источник информации и не пытается заменить собой остальные.

Третий вывод связан с энтропией. Её очень легко интерпретировать неправильно. Если энтропия распределения рекомендаций высокая, это означает, что алгоритм не показывает всем пользователям одно и то же небольшое множество товаров. Но из этого ещё не следует, что конкретный пользователь получил полезную для него рекомендацию.

Обратная ситуация тоже возможна. Низкая энтропия списка для одного клиента означает лишь то, что список получился узким и сосредоточенным на небольшом числе категорий. Это не доказывает персонализацию: рекомендации могут быть очень однообразными и при этом совершенно не соответствовать потребностям клиента. Поэтому энтропию нельзя рассматривать отдельно от метрик, оценивающих качество ранжирования. Такой подход соответствует современной практике beyond-accuracy-оценки рекомендательных систем [14; 16].

Полученные результаты в целом соответствуют направлению современных исследований B2B-рекомендаций, но решают другую задачу. Например, в исследовании IndiaMART основное внимание уделялось сопоставлению запросов на коммерческие предложения с продавцами и работе с дисбалансом классов [12]. В корейском исследовании модель KGAT применялась для поиска потенциальных клиентов [10]. В работах по российским регулируемым закупкам оценивалась вероятность успешного завершения закупочной процедуры [3].

В настоящем исследовании задача существенно уже: требуется предсказать следующую товарную категорию, которую может приобрести клиент. Поэтому прямое сравнение численных результатов между этими исследованиями было бы некорректным. Если модели решают разные задачи, используют разные данные и измеряются разными метриками, одинаковое числовое значение показателя ещё не означает одинаковое качество.

Отдельно важно различать качество модели и экономический эффект. Рост офлайн-метрики ещё не означает рост выручки или прибыли. Допустим, модель правильно угадала товар, который заинтересовал клиента. Эта рекомендация всё равно бесполезна для бизнеса, если товара нет на складе, клиенту недоступна нужная цена, нарушается минимальная партия, упаковка имеет неподходящую кратность, превышен кредитный лимит или продажа даёт слишком низкую маржу.

Поэтому промышленная рекомендательная система должна состоять как минимум из двух уровней. Первый уровень оценивает статистическую привлекательность товара для клиента. Второй проверяет, можно ли этот товар реально предложить с учётом коммерческих ограничений. В такой архитектуре разработанный метод представляет собой ядро ранжирования, а не самостоятельный автомат продаж.

Дальнейшее развитие модели целесообразно вести по трём направлениям. Первое – замена категориального ранжирования на SKU-уровень с учётом атрибутов и товарного графа; мультимодальный мэтчинг может использоваться для нормализации каталога [2]. Второе – отдельные временные модели для replenishment и риска оттока, поскольку они не сводятся к ранжированию следующей категории. Третье – иерархическое или нейросетевое повторное ранжирование после накопления логов показов.

Заключение

В работе разработана и вычислительно проверена кластерно-гибридная модель товарных рекомендаций для B2B-платформы. Модель объединяет item-based коллаборативную фильтрацию, профиль K-Prototypes, отраслевой приор и штраф за глобальную популярность. Такая архитектура соответствует специфике межфирменной торговли, где индивидуальная история редка, а профиль организации и её отрасль несут самостоятельную информацию.

На синтетической панели из 2000 клиентов и 25 000 транзакций гибрид достиг NDCG@5 = 0,4650 и превзошёл ранжирование по популярности на

14,3%, кластерный метод на 15,4% и item-based CF на 34,2%. Разность с сильнейшим базовым методом равна 0,0582, 95%-й бутстреп-интервал составляет [0,0470; 0,0695]. Одновременно нормированная энтропия экспозиций увеличилась с 0,6477 до 0,9067, а доля непопулярных категорий составила 0,2447.

Полученные данные уточняют исходную научную гипотезу. Положительный эффект даёт не кластеризация как изолированный метод, а её использование в связке с индивидуальным и отраслевым сигналами. Авторский вклад состоит в этой формализации, в явном ограничении роли энтропии и в воспроизводимой процедуре временной оценки.

Вместе с тем исследование доказывает только внутреннюю работоспособность метода. Синтетическая панель не может подтвердить рост выручки, удержания или сокращение трудозатрат. Для внешней валидизации необходимы анонимизированные данные реальной B2B-платформы, проверка коммерческой допустимости и кластерный A/B-тест. Эта последовательность образует логическое продолжение работы.

Литература

1. Data Insight, B2B-РТС. Отраслевой отчёт по B2B и B2G онлайн-продажам и закупкам: текущее состояние, объёмы рынка, тренды и прогноз 2025–2030. 2025. URL: https://datainsight.ru/DI_b2b_b2g_eCommerce (дата обращения: 13.08.2026).
2. Варнухов А. Ю., Назаров Д. М. Мэтчинг товаров на маркетплейсах: мультимодальная модель на основе архитектуры трансформера // Бизнес-информатика. 2025. Т. 19, № 2. С. 7–24. DOI: 10.17323/2587-814X.2025.2.7.24.
3. Денисова А. И., Созаева Д. А., Гончар К. В. Разработка рекомендательных систем для повышения эффективности регулируемых закупок в электроэнергетике // Бизнес-информатика. 2025. Т. 19, № 2. С. 25–40. DOI: 10.17323/2587-814X.2025.2.25.40.
4. Краснов Ф. В. Анализ эффективности Learning-to-Rank в каталогах B2B для e-commerce: цифровой эксперимент и оценка влияния на конверсию // Бизнес-информатика. 2026. Т. 20, № 1. С. 54–66. DOI: 10.17323/2587-814X.2026.1.54.66.
5. Соколов А. В., Сычев И. А., Соколова О. Л. и др. Теоретические и практические аспекты построения рекомендательных моделей: типология, архитектура и направления проектирования // Вестник Пермского университета. Математика. Механика. Информатика. 2024. № 3(66). С. 64–77. DOI: 10.17072/1993-0550-2024-3-64-77.
6. Стребкова Л. Н. Особенности развития электронной торговли в России: B2C и B2B // Вестник НГУЭУ. 2025. № 2. С. 132–145. DOI: 10.34020/2073-6495-2025-2-132-145.
7. Adomavicius G., Tuzhilin A. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions // IEEE Transactions on Knowledge and Data Engineering. 2005. Vol. 17, no. 6. P. 734–749. DOI: 10.1109/TKDE.2005.99.
8. Burke R. Hybrid Recommender Systems: Survey and Experiments // User Modeling and User-Adapted Interaction. 2002. Vol. 12. P. 331–370. DOI: 10.1023/A:1021240730564.
9. Carrizosa E., Kurishchenko K., Marín A., Romero Morales D. Interpreting Clusters via Prototype Optimization // Omega. 2022. Vol. 107. Article 102543. DOI: 10.1016/j.omega.2021.102543.
10. Cho G., Shim P.-S., Kim J. Explainable B2B Recommender System for Potential Customer Prediction Using KGAT // Electronics. 2023. Vol. 12, no. 17. Article 3536. DOI: 10.3390/electronics12173536.
11. Davies D. L., Bouldin D. W. A Cluster Separation Measure // IEEE Transactions on Pattern Analysis and Machine Intelligence. 1979. Vol. 1, no. 2. P. 224–227. DOI: 10.1109/TPAMI.1979.4766909.
12. Gaur V., Liu X., Dhaliwal A., Jain A., Manchanda G. Personalized Recommender System Design for IndiaMART's Online B2B Platform // Analytics Magazine. 2021. DOI: 10.1287/LYTX.2021.04.26n.
13. Huang Z. Extensions to the K-Means Algorithm for Clustering Large Data Sets with Categorical Values // Data Mining and Knowledge Discovery. 1998. Vol. 2, no. 3. P. 283–304. DOI: 10.1023/A:1009769707641.
14. Kaminskas M., Bridge D. Diversity, Serendipity, Novelty, and Coverage: A Survey and Empirical Analysis of Beyond-Accuracy Objectives in Recommender Systems // ACM Transactions on Interactive Intelligent Systems. 2016. Vol. 7, no. 1. Article 2. DOI: 10.1145/2926720.
15. Koren Y., Bell R., Volinsky C. Matrix Factorization Techniques for Recommender Systems // Computer. 2009. Vol. 42, no. 8. P. 42–49. URL: https://gwern.net/doc/ai/tabular/2009-koren.pdf (access date: 18 August 2026)
16. Lin S., Gao C., Chen J. et al. How Do Recommendation Models Amplify Popularity Bias? An Analysis from the Spectral Perspective // Proceedings of the 18th ACM International Conference on Web Search and Data Mining. 2025. P. 659–668. DOI: 10.1145/3701551.3703579.
17. Rendle S., Freudenthaler C., Gantner Z., Schmidt-Thieme L. BPR: Bayesian Personalized Ranking from Implicit Feedback // Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence. 2009. P. 452–461.
18. Rousseeuw P. J. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis // Journal of Computational and Applied Mathematics. 1987. Vol. 20. P. 53–65. DOI: 10.1016/0377-0427(87)90125-7.
19. Shambour Q., Lu J. An Effective Recommender System by Unifying User and Item Trust Information for B2B Applications // Journal of Computer and System Sciences. 2015. Vol. 81. P. 1110–1126. DOI: 10.1016/j.jcss.2014.12.029.
20. Shannon, C. E. A Mathematical Theory of Communication // Bell System Technical Journal. 1948. Vol. 27, no. 3. P. 379–423; no. 4. P. 623–656.
21. Zhang X., Wang H. Study on Recommender Systems for Business-to-Business Electronic Commerce // Communications of the IIMA. 2005. Vol. 5, iss. 4. P. 53–61. DOI: 10.58729/1941-6687.1282.

Cluster-Hybrid Product Recommendation Model for Corporate Online Procurement

Moiseev N.A., Odoevsky I.I.
Plekhanov Russian University of Economics

JEL classification: B00, D20, E22, E44, L23, L51, L52, M11, M20, M30, Z33

The object of the study is recommender systems used by corporate electronic trading platforms. The subject is the ranking of product categories under sparse purchasing histories and heterogeneous account profiles. The purpose is to develop and computationally test a cluster-hybrid model that combines item-based collaborative filtering, K-Prototypes group profiles, an industry prior, and a global-popularity penalty. The experiment uses a fully reproducible synthetic panel of 2,000 corporate accounts and 25,000 transactions across 12 product categories. Each account's events are divided chronologically into training, validation, and test subsets in a 60/20/20 ratio. Performance is measured with Precision@5, Recall@5, NDCG@5, catalog coverage, long-tail exposure share, and normalized entropy of aggregate exposure. The proposed model achieves NDCG@5 of 0.4650, compared with 0.4068 for popularity ranking, 0.4029 for the cluster-only method, and 0.3464 for item-based collaborative filtering. The gain over the strongest baseline is 0.0582, with a 95% bootstrap interval of [0.0470; 0.0695]. Normalized exposure entropy rises from 0.6477 to 0.9067, while the long-tail share increases from zero to 0.2447. The scientific contribution is a unified ranking framework in which segmentation complements rather than replaces personalization and is combined with organizational context and explicit popularity-bias correction. The findings demonstrate internal validity in a controlled setting but do not constitute evidence of production impact without testing on real transaction data.

Keywords: B2B e-commerce, K-Prototypes, implicit feedback, popularity bias, exposure entropy, synthetic data, ranking quality.

References

1. Data Insight, B2B-RTS (2025). Industry report on B2B and B2G online sales and procurement: Current state, market size, trends and forecast 2025–2030. Available at: https://datainsight.ru/DI_b2b_b2g_eCommerce (accessed 13 August 2026).
2. Varnukhov A. Yu., Nazarov D. M. (2025). Product matching in digital marketplaces: Multimodal model based on the transformer architecture. Business Informatics, vol. 19, no. 2, pp. 7–24. DOI: 10.17323/2587-814X.2025.2.7.24.
3. Denisova A. I., Sozaeva D. A., Gonchar K. V. (2025). Development of recommendation systems to improve the efficiency of regulated procurement in the electric power industry. Business Informatics, vol. 19, no. 2, pp. 25–40. DOI: 10.17323/2587-814X.2025.2.25.40.
4. Krasnov F. V. (2026). Learning-to-Rank in B2B e-commerce catalogs: A digital experiment and conversion analysis. Business Informatics, vol. 20, no. 1, pp. 54–66. DOI: 10.17323/2587-814X.2026.1.54.66.
5. Sokolov A. V., Sychev I. A., Sokolova O. L. et al. (2024). Theoretical and practical aspects of building recommendation models: Typology, architecture and design directions. Bulletin of Perm University. Mathematics. Mechanics. Computer Science, no. 3(66), pp. 64–77. DOI: 10.17072/1993-0550-2024-3-64-77.