

DOI: 10.17323/2587-814X.2025.2.7.24

Мэтчинг товаров на маркетплейсах: мультимодальная модель на основе архитектуры трансформера

А.Ю. Варнухов

E-mail: varnuhov_ayu@usue.ru

Д.М. Назаров 

E-mail: slup2005@mail.ru

Уральский государственный экономический университет, Екатеринбург, Россия

Аннотация

В работе рассматривается проблема интеллектуального мэтчинга товаров на маркетплейсах в рамках которого необходимо выполнять оценку схожести различных записей, которые описывают продукты, но могут отличаться по формату, содержанию или объёму мультимодальных данных. Тематика научного поиска находится на пересечении методов решения задач entity resolution (ER) — сопоставления записей и мультимодального анализа данных. Она является крайне актуальной в условиях быстрорастущей платформенной экономики при кратно увеличивающемся объеме рынка электронной коммерции. Основная цель исследования — разработать и протестировать мультимодальную интеллектуальную модель на базе архитектуры трансформера для повышения точности и робастности мэтчинга товаров на маркетплейсах. Разработанная авторами модель, интегрирующая текстовые, визуальные и табличные признаки, позволит эффективнее идентифицировать схожие товары, проводить поиск конкурентных предложений, обнаруживать дубликаты, а также выполнять кластеризацию и сегментацию товаров. Предложенный подход опирается на механизм внимания (self-attention), позволяющий моделировать контекстуально-семантические взаимосвязи в данных различной природы. Для извлечения векторного представления текстовых описаний используются языковые модели, в частности архитектура Sentence-BERT, для графической компоненты Vision Transformer, а табличные данные обрабатываются с применением специализированных механизмов обучения на структурированных данных TabTransformer. Проведенный эксперимент продемонстрировал, что разработанная мультимодальная модель эффективно справляется с задачей мэтчинга в условиях значительной вариативности товарных позиций и неоднородности данных. Кроме того, полученные данные позволяют утверждать, что модель может быть успешно адаптирована для использования в рам-

* Статья опубликована при поддержке Программы НИУ ВШЭ «Университетское партнерство»

ках других категорий товаров. Полученные результаты подтверждают эффективность и целесообразность использования мультимодального подхода для реализации мэтчинга товаров на маркетплейсах, позволяя участникам рынка электронной коммерции существенно улучшить качество управления товарными запасами, повысить эффективность ценообразования и укрепить свои конкурентные преимущества.

Ключевые слова: маркетплейсы, контекстно-семантическая идентификация, поиск конкурентных предложений, мэтчинг, машинное обучение, глубокое обучение, архитектура трансформера, интеллектуальный анализ данных

Цитирование: 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

Введение

Стремительное развитие платформенной экономики в последние десятилетия вызвано появлением и повсеместным распространением цифровых платформ, которые выступают в качестве медиатора между покупателями и продавцами. Маркетплейсы, как цифровые платформы, позволяют продавцам получать доступ к широкой аудитории, а покупателям — к удобным сервисам поиска и сравнения товарных предложений от различных продавцов. При этом управление данными о товарах и их сопоставление (мэтчинг) становятся сложной научной проблемой, от решения которой напрямую зависят прямые и косвенные показатели экономической деятельности всех участников платформ электронной коммерции. К прямым можно отнести показатели, влияющие на выбор модели ценообразования, а к косвенным — эффективность продвижения продукции, точность рекомендаций и удовлетворённость потребителей. При этом совершенно очевидно, что прямые и косвенные показатели связаны каузально. Актуальность темы исследования обусловлена тем, что с ростом объёмов рынков электронной коммерции и расширением номенклатурных позиций товаров на цифровых платформах возникает потребность в использовании более продвинутых математических и инструментальных методов в экономике, способных обеспечить автоматизированный, а скорее даже интеллектуальный и надёжный мэтчинг товаров в больших потоках разнородной информации на маркетплейсах.

С научной точки зрения проблема мэтчинга товаров (идентификации записей о товарах) близка к задачам entity resolution (далее ER) — сопоставления сущностей с учетом возможности их дублирования, однако в контексте маркетплейсов, расширяется за счёт использования мультимодальных данных при описании товаров. Действительно, описания товаров на маркетплейсах обычно включают не только текстовую информацию, но также изображения и их характеристики, выраженные в табличной форме.

Ниже приводится краткий обзор литературы, разбитый на два логических блока — персонализация с учетом специфики e-commerce и ER в контексте алгоритмических подходов к мэтчингу.

Ряд работ, которые легли в основу формирования научной базы данного исследования, фокусируются на изучении поведения пользователей, создании рекомендательных систем и оценке особенностей товарных категорий в электронной коммерции (Н. Angermann, М. Mao, F.T. Abdul Hussien, A. Fletcher, P. Ristoski, M. Cheung и др.). Большинство ученых указывают на важность использования интегрированного подхода при обработке текстовых описаний, изображений и метаданных для повышения точности рекомендаций и выявления скрытых взаимосвязей в больших массивах данных (big data). Анализ этих исследований в контексте мэтчинга показывает необходимость применения технологий обработки мультимодальных данных, чтобы в конечном итоге корректно сопоставлять не только названия и характеристики, но и визуальные атрибуты товаров.

Другим направлением исследований в рамках тематики статьи являются классические задачи поиска дубликатов записей, сопоставления сущностей и разработки алгоритмов для точного объединения данных (W.W. Cohen, S.S. Aalen, J. Devlin, A.K. Elmagarmid, N. Reimers, J. Wang, H. Köpcke и др.). В работах этих авторов рассматриваются различные метрики сходства, включая расстояние Левенштейна и аналоги, языковые модели и методы обучения по прецедентам, позволяющие конструировать паттерны для выявления потенциальных совпадений. Кроме того, нередко поднимается вопрос о масштабируемости решений — от разработки пакетных программных продуктов до высоконагруженных систем, обрабатывающих динамические потоки данных в режиме реального времени. Стремительное развитие архитектур глубокого обучения, в том числе на основе трансформеров, обуславливает переход от автоматизированного поиска признаков (далее features) к интеллектуальным методам, выявляющим сложные паттерны. Такие подходы особенно актуальны для маркетплейсов, где критически важными факторами являются быстрота и надёжность обработки больших массивов данных, учитывающих их различные экономические характеристики, в том числе и динамику цен.

Таким образом, существующие исследования подтверждают важность применения продвинутых экономико-математических методов при разработке систем мэтчинга товаров на маркетплейсах на основе обработки мультимодальных данных. Однако, ряд вопросов в рамках проведенного обзора литературы остаётся открытыми: как оптимально объединить признаки, какие подходы наиболее эффективны для анализа мультимодальных данных и каким образом адаптировать разрабатываемые модели к появлению новых товаров и источников информации. В рамках данного исследования предлагается мультимодальная модель, основанная на архитектуре трансформера, которая способна обрабатывать данные из различных источников и позволяет последовательно и эффективно интегрировать разные модальности благодаря механизму внимания и многоуровневым представлениям для более точного сопоставления товарных позиций. Предложенный подход вносит вклад в область математических и инструментальных методов в экономике, направленных на создание устойчивых систем анализа больших данных, а также поддерживает масштабируемость и учитывает высокую динамичность платформ электронной коммерции.

Научный вклад работы состоит в создании и валидации подхода, который позволяет интегрировать одновременно текстовые, визуальные и табличные данные в единой модульной системе на основе трансформер-архитектуры.

Цель исследования — разработать и апробировать такую модель, чтобы обеспечить высокую точность и робастность мэтчинга товаров на больших выборках товарных позиций, характерных для современных маркетплейсов.

Статья состоит из четырех разделов. Первый раздел посвящен изучению потенциала и возможностей применения мэтчинга товаров на маркетплейсах. Во втором разделе проводится анализ существующих подходов и делается постановка задачи мэтчинга на маркетплейсах. В третьем разделе разработана мультимодальная модель мэтчинга товаров на маркетплейсах (МММР). В четвертом разделе исследования обсуждается применение модели для оценки схожести товаров на маркетплейсе Wildberries.

1. Потенциал и возможности применения мэтчинга товаров на маркетплейсах

Под мэтчингом товаров (product matching) понимается процесс сопоставления (связывания) товаров, который позволяет определять какие позиции являются одинаковыми или аналогичными по своей сути, чтобы гарантировать, что один и тот же продукт будет правильно распознан даже если он появится под разными названиями, описаниями или идентификаторами. Так, например, при наличии эквивалентных моделей смартфона от одного бренда, представленного в разных магазинах на маркетплейсах, его характеристики, фото и даже формулировка наименования могут существенно отличаться.

Мэтчинг товаров является фундаментальным компонентом, который активно применяется в различных современных цифровых экосистемах, включая маркетплейсы, платформы объявлений, сервисы электронной коммерции и другие онлайн-площадки. С точки зрения покупателей, мэтчинг товаров позволяет существенно повысить удобство совершения покупок поскольку обеспечивает возможность простого и быстрого поиска аналогичных товарных предложений от различных продавцов [1]. Например, внедрив систему подбора товаров, которая

объединит схожие предложения в единую и хорошо структурированную коллекцию однотипных объявлений можно избавить покупателей от необходимости перебирать многочисленные варианты в рамках одной товарной категории. Подобная система будет ограждать покупателей от просмотра дублирующих или вводящих в заблуждение предложений, которые превращают подбор товаров в запутанный и отнимающий продолжительное время процесс. Вместо этого покупатели получают возможность быстро сравнивать цены, отзывы и рейтинг продавцов, что, в конечном итоге, приведет к повышению релевантности поиска и общей удовлетворенности от совершения покупок, а также будет способствовать принятию более взвешенных и обоснованных решений. Кроме того, мэтчинг товаров может применяться для создания персонализированных рекомендаций на основе анализа поведения потребителей [2], их предпочтений и прошлых покупок с целью оказания помощи при выборе наиболее подходящих товаров, повышая вовлеченность пользователей и общую ценность платформы.

С точки зрения продавцов, мэтчинг товаров помогает анализировать конкурентные предложения [3], позволяя корректировать и выстраивать свои ценовые стратегии на основе рыночных тенденций в режиме реального времени [4]. Цифровые торговые площадки аккумулируют огромные объемы данных о товарах, однако без должной обработки и надлежащей структурной организации большая часть этой информации остается фрагментированной, противоречивой и сложной для эффективного анализа. Например, многие продавцы экспонируют один и тот же товар на платформах с некоторыми отличиями в названии, фотографиях, описании или характеристиках. Без применения мэтчинга товаров продавцам приходится вручную отслеживать и сравнить тысячи объявлений, что является весьма энергоемким и затратным процессом, который чреват ошибками. Автоматизированная идентификация и увязка схожих товаров предоставляет возможность объединить разрозненную информацию в единое согласованное хранилище данных, которое будет являться ценным источником сведений для дальнейшего анализа и принятия решений на основе актуальной рыночной ситуации. Обладание таким источником сведений создает условия для применения моделей алгоритмического ценообразования, которые основаны на оценке рыночной ситуации поскольку позволяют проводить оперативный мониторинг цен конкурентов и

изменений рыночного спроса, открывая возможность для применения стратегий динамического ценообразования. Благодаря этому продавцы могут адаптировать свои предложения в режиме реального времени в соответствии с поведением потребителей и текущей рыночной ситуацией, гарантируя, что цены на товары остаются конкурентоспособными, привлекая покупателей без ущерба для своей маржинальности. Подобная гибкость позволяет выявлять зарождающиеся тренды, активно реагировать на сезонные колебания, рекламные активности и экономические изменения, что в конечном итоге будет способствовать долгосрочному и устойчивому развитию в условиях высококонкурентного рынка цифровых платформ. Также мэтчинг товаров играет важную роль в проведении рекламных и маркетинговых кампаний. Одним из основных преимуществ использования мэтчинга товаров в этом контексте является возможность оптимизировать расходы на рекламу. При правильной идентификации аналогичных товаров, предлагаемых другими продавцами, можно избежать излишней конкуренции, что позволит исключить напрасные расходы и перераспределить бюджет в более выгодные ниши. Помимо рекламных активностей расширяются возможности кросс-сейла. Аккуратно связывая взаимодополняющие, часто покупаемые или другие полезные товары друг с другом продавцы могут формировать привлекательные предложения.

Еще одним важным направлением применения алгоритмов мэтчинга является их роль в борьбе с мошенничеством и выявлении контрафактной продукции для оказания содействия операторам платформы в поддержании доверия, безопасности и целостности экосистем [5]. Вместе с быстрым развитием и ростом влияния площадок электронной коммерции на общественную жизнь все чаще возникают инциденты, связанные с мошенническими объявлениями, контрафактными и умышленно вводящими в заблуждение товарами, что указывает на серьезный вызов для отрасли. Подобные противоправные действия злоумышленников наносят ущерб не только конечным потребителям, но также подрывают общее доверие к цифровым площадкам и бизнесу законопослушных продавцов [6]. Необходимость обнаружения контрафакта особенно актуальна для многих товарных категорий, среди которых можно привести в пример: электронику, фармацевтику и косметику. В этих категориях контрафактные товары могут нести не только финансо-

вые риски, но также представлять непосредственную угрозу здоровью и безопасности потребителей. В этом контексте алгоритмы мэтчинга являются важным инструментом, который позволяет предотвращать появление на площадке предложений и объявлений от неавторизованных посредников или продажи некачественной контрафактной продукции под видом известных торговых марок.

Исходя из вышеизложенного допустимо сделать вывод, что потенциал и возможности применения мэтчинга товаров выходят далеко за рамки односложного инструмента для решения единственной задачи. Можно утверждать, что это основополагающая технология, которая позволяет оптимизировать множество различных бизнес-процессов, вести борьбу с мошенничеством и содействует повышению эффективности процесса ценообразования, формируя более прозрачную и удобную экосистему для участников рынка электронной коммерции — потребителей и продавцов.

2. Анализ существующих подходов и постановка задачи мэтчинга на маркетплейсах

2.1. Мэтчинг товаров на основе сходства значений атрибутов и теоретико-множественные методы

Сопоставление атрибутов товаров на основе сходства значений можно отнести к базовым подходам мэтчинга, который основывается на сравнении текстовых и числовых полей в различных характеристиках с целью определения насколько точно они совпадают между собой [7]. Текстовые атрибуты оцениваются с помощью таких метрик как расстояние Левенштейна, Джаро-Винклера или TF-IDF (Term Frequency — Inverse Document Frequency), которые позволяют определить, насколько близко два фрагмента текста похожи друг на друга с учетом имеющихся различий в написании [8]. Числовые атрибуты (цена, вес, габаритные размеры и так далее) обычно оцениваются путем вычисления абсолютного или относительного отклонения. После подготовки и вычисления индивидуальных метрик сходства для каждого атрибута эти метрики объединяются, например с помощью взвешенной суммы, для получения общей оценки соответствия сравниваемых товаров. Если полученный показатель преодолевает установленное пороговое значение, то пара сравниваемых товаров

может считаться похожей. Кроме того, в теоретико-множественном подходе каждая карточка товара моделируется как множество атомарных элементов (признаков, n-грамм, токенов и так далее). Сходство между объектами определяется с помощью таких классических метрик, как коэффициенты Жаккара, Сёренсена—Дайса или Шимкевича—Симпсона. За счёт простоты вычислений данный подход позволяет обеспечить высокую скорость и простую масштабируемость, однако игнорирует порядок и контекст, что существенно ограничивает его точность. К ключевым преимуществам данного подхода можно отнести то, что он относительно прост и прямолинеен, характеризуется высокой прозрачностью и может быть легко реализован на практике. Среди существенных недостатков можно выделить трудности при работе с большими и разнообразными товарными категориями, проблемы с некачественными или зашумленными данными, а также крайне ограниченные возможности по выделению семантических значений.

2.2. Мэтчинг товаров на основе экспертной системы правил

Логическим развитием подхода на основе сравнения значений атрибутов стала разработка подхода на основе системы правил [9, 10]. Мэтчинг, основанный на системе правил строится с помощью определяемых экспертами логических конструкций, которые описывают функционирование конкретной предметной области и позволяют определить степень схожести оцениваемых товаров. Каждое правило обычно оценивает подмножество четко заданных атрибутов и использует простые логические условия или пороговые значения (например, «Если название бренда идентично, а расстояние Левенштейна модели меньше двух, то такие два товара можно считать схожими»). Поскольку данный подход основан на системе правил, которые напрямую кодируют экспертные знания, их можно быстро и легко интерпретировать. Однако при работе с обширными и комплексными категориями система правил получается весьма громоздкой и сложной в обслуживании, а также нуждается в постоянном ручном обновлении и расширении. Кроме того, поскольку каждое правило должно быть определено экспертами вручную, эта система подвержена ошибкам и в конечном итоге оказывается недостаточно адаптированной и излишне перегруженной.

2.3. Мэтчинг товаров на основе таксономий и онтологий

Эти методы опираются на глубокий контекстуально-семантический анализ сведений о предметной области и применяют структурированное представление для отображения взаимосвязей товаров и их атрибутов в виде графа знаний [11, 12]. Взаимосвязи могут указывать на общие характеристики (например, на определенный бренд или семейство продуктов), иерархическое наследование или дескриптивные взаимоотношения. Моделирование данных с помощью графа зависимостей открывает возможность более глубокого исследования доступной информации, позволяя учитывать контекст, внутренние отношения между сущностями и скрытые паттерны. Вместо того, чтобы полагаться исключительно на попарное сравнение атрибутов, данный подход позволяет выявлять комплексные зависимости, оперировать сложными логическими связями и оценивать структурные характеристики в совокупности. Кроме того, построенные онтологии и таксономии могут внести весомый вклад при проведении стандартизации или согласовании разрозненных сведений, что имеет важное значение при интеграции данных из нескольких источников [13]. Среди ключевых преимуществ данного подхода можно выделить его холистический взгляд, который позволяет обнаруживать кластеры и связанные компоненты эквивалентных товаров путем анализа целых подграфов, иногда выявляя совпадения, пропущенные при использовании более простых подходов. Однако, построение и поддержание всеобъемлющего графа знаний, как правило, является трудоемким и ресурсоемким процессом, который нуждается в регулярном обновлении при расширении ассортимента товаров или появлении новых источников данных. Также стоит отметить, что неполные или противоречивые таксономии и онтологии могут критически влиять на точность мэтчинга, нивелируя все преимущества данного подхода.

2.4. Мэтчинг товаров на основе машинного обучения

Подходы на основе машинного обучения и глубоких нейросетей рассматривают задачу мэтчинга как задачу оптимизации целевого функционала [14, 15]. В этой формулировке обученная модель оценивает множество атрибутов и определяет со-

ответствуют ли они аналогичному товару. Методы классического машинного обучения начинаются с выбора и инжиниринга коллекции признаков, которые позволяют репрезентативно представить товар. Затем эти подготовленные признаки передаются в некоторый бинарный классификатор (логистическая регрессия, случайный лес, SVM и так далее), который тренируется отличать похожие товары на основе зависимостей в исходном пространстве признаков. Использование таких методов требует наличия предварительно размеченного набора данных с указанием метки целевого класса, который позволяет модели определить как сходства или различия в наборе признаков влияют на вероятность совпадения товаров. Методы классического машинного обучения могут обеспечить высокую точность предсказаний, однако нуждаются в разметке, сильных знаниях предметной области и тщательном инжиниринге исходных признаков, что вызывает необходимость постоянной адаптации при изменениях в данных или их распределении. Методы глубокого обучения позволяют значительно сократить или полностью устранить затраты, которые связаны с инжинирингом признаков за счет автоматического выявления паттернов в данных. В качестве базовых моделей могут применяться рекуррентные нейронные сети (RNN), сети с долгой краткосрочной памятью (LSTM), а также сверточные нейронные сети (CNN). После обработки такой моделью исходных признаков формируется эмбединг, представляющий собой характеристический вектор, который затем используется для оценки схожести товаров. К основным преимуществами моделей на основе глубокого обучения можно отнести их более высокую точность обработки зашумленных или разнородных данных, отсутствие необходимости тщательного инжиниринга признаков, а также повышенную устойчивость при обработке ранее неизвестных данных. Однако они сталкиваются с рядом проблем, которые связаны с выделением и пониманием семантического значения, а также требуют наличия предварительно размеченных данных.

2.5. Постановка задачи мэтчинга товаров на маркетплейсах

Исходя из проведенного анализа и особенностей функционирования маркетплейсов становится понятно, что к типичным сложностям мэтчинга можно отнести: неоднозначность описаний и исполь-

зование маркетинговых терминов, дублирование и пропуски в данных, мультиформатность (текстовые описания, изображения, табличные параметры) и различный уровень детализации. Все эти факторы повышают риски ошибочного сопоставления товаров или, наоборот, пропуска потенциальных совпадений. Таким образом, требуется разработать мультимодальную модель, способную обрабатывать как визуальную компоненту, так и характеристики товара с учетом их семантического значения. Такая постановка задачи связана с разнородным характером данных и высокой вариативностью цифровой среды. Использование только характеристик часто оказывается недостаточным и может привести к ошибке мэтчинга, особенно если речь идет о товарах, отличительные черты которых заметны на изображениях. Аналогично, изображения сами по себе могут быть неоднозначными или выглядеть практически идентично. Таким образом, за счёт интеграции этих компонентов в единую систему, модель будет более эффективно и комплексно выполнять мэтчинг.

3. Мультимодальная модель мэтчинга товаров на маркетплейсах (МММР)

Пусть задано множество информационных карточек $\Omega = \{p_1, p_2, p_3, \dots, p_n\}$, содержащее сведения о каждом товаре. Требуется разработать модель Ψ (1), которая отображает множество Ω в пространство $\mathbb{R}^d$ таким образом, чтобы данное векторное представление позволяло оценить степень подобия между любыми парами товаров из Ω согласно (2), где S мера сходства:

$$\Psi(p_i) = (x_1, x_2, x_3, \dots, x_d); p_i \in \Omega, \quad (1)$$

$$S: \mathbb{R}^d \times \mathbb{R}^d \rightarrow [0, 1]. \quad (2)$$

Так как модель должна обрабатывать несколько модальностей, то ее концептуальное представление можно разложить на составляющие компоненты согласно (3):

$$\Psi(p_i) = \text{Concat}(\psi_{\text{title}}(p_i), \psi_{\text{image}}(p_i), \psi_{\text{metadata}}(p_i)); p_i \in \Omega, \quad (3)$$

где ψ_{title} кодирует в векторное представление название товара;

ψ_{image} кодирует в векторное представление изображение товара;

ψ_{metadata} кодирует в векторное представление характеристики товара.

Поскольку необходимо учитывать контекст и семантические значения, то в настоящей работе предлагается использовать в качестве основы архитектуру трансформера, предложенную в [16]. Известно, что эта архитектура позволяет представить текстовые данные в векторном виде и, благодаря механизму внимания в сочетании с глубокими нейронными сетями, эффективно улавливает семантические и контекстуальные особенности [17]. В отличие от традиционных моделей генерации эмбеддингов, таких как Word2Vec [18] и GloVe [19], назначающих каждому слову фиксированный вектор, трансформер создает динамические, контекстно-связанные эмбеддинги. Это означает, что одно и то же слово может иметь разные представления в зависимости от контекста его использования в предложении. Учитывая, что в настоящей работе решается задача векторного представления, то интерес представляет исключительно блок кодировщика. Архитектура кодировщика включает три ключевых элемента: многослойный механизм внимания (self-attention), слой остаточного соединения и нормализации, а также нейронную сеть прямого распространения. Рассмотрим основные преобразования.

Пусть задано множество входных токенов $t = \{t_1, t_2, t_3, \dots, t_n\}$, каждому из которых сопоставим собственный эмбеддинг как указано в (4):

$$\text{Embed}: t_n \rightarrow e_i \in \mathbb{R}^{d_{\text{model}}}, \quad (4)$$

где d_{model} соответствует размерности эмбеддингов модели.

Для учета порядка следования элементов используется позиционная кодировка вида (5, 6):

$$PE(pos, 2i) = \sin\left(\frac{pos}{10000^{2i/d_{\text{model}}}}\right), \quad (5)$$

$$PE(pos, 2i+1) = \cos\left(\frac{pos}{10000^{2i/d_{\text{model}}}}\right), \quad (6)$$

где pos — позиция в последовательности;

i — размерность.

Затем на вход первого блока кодировщика подается (7):

$$X = (e_1 + \rho e_1, e_2 + \rho e_2, \dots, e_n + \rho e_n) \in \mathbb{R}^{n \times d_{\text{model}}}. \quad (7)$$

Каждый блок кодировщика содержит элемент механизма внимания, использующий матрицы за-

просов (Q), ключей (K) и значений (V), которые определяются согласно (8):

$$Q = XW^Q, K = XW^K, V = XW^V; \\ W^Q, W^K, W^V \in \mathbb{R}^{d_{model} \times d_k}, \quad (8)$$

где W^Q , W^K и W^V – обучаемые параметры модели;

d_k – обычно определяется как отношение d_{model} к количеству слоев механизма внимания.

Оценка значимости на основе механизма внимания вычисляется согласно (9):

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V. \quad (9)$$

В архитектуре трансформера применяется множество слоев внимания, которые конкатенируются между собой согласно (10) и (11):

$$\text{head}_i = \text{Attention}(Q_i, K_i, V_i), \quad (10)$$

$$\text{MultiHead}(X) = \\ = \text{Concat}(\text{head}_1, \text{head}_2, \text{head}_3, \dots, \text{head}_n)W^O, \quad (11)$$

где матрица W^O также является обучаемым параметром модели.

Важной частью архитектуры трансформера является слой остаточного соединения [20] и нормализации [21], применяемый после механизма внимания согласно (12):

$$X_{attm} = \text{LayerNorm}(X + \text{MultiHead}(X)). \quad (12)$$

Каждый блок кодировщика также содержит нейронную сеть прямого распространения обычно в форме двухслойного MLP (13) с нелинейностью вида ReLU (или GeLU):

$$\text{FFN}(X_{attm}) = \max(0, X_{attm} W_1 + b_1) W_2 + b_2, \quad (13)$$

после которой повторно применяется слой остаточного соединения и нормализации. Полученный выход одного блока кодировщика последовательно передается через N аналогичных кодировщиков, формируя в результате их выполнения итоговый результат. Концептуальное представление архитектуры кодировщика показано на *рисунке 1*.

Задача определения степени схожести названия товара требует надежного представления семантики предложений, которое позволяет эффективно и точно сопоставлять текстовое содержимое. Традиционные модели на основе архитектуры трансформера, такие как BERT (Bidirectional Encoder Representations from Transformers) [22], продемонстрировали значительный успех в различных проблемах обработки естественного языка (NLP). Однако BERT оптимизирован для решения задач на уровне токенов и не создает целостных эмбеддингов фиксированного размера на уровне предложений. Для преодоления этого ограничения предлагается использовать Sentence-BERT [23], который специально разработан для генерации эмбеддингов на уровне предложений и использует структуру сиамской нейронной сети. В отличие от BERT, который обрабатывает каждую пару отдельно, Sentence-BERT выполняет кодирование в плотный вектор фиксированного размера, что позволяет эффективно вычислять косинусное сходство предложений в векторном пространстве. Кроме того, Sentence-BERT может обучаться с использованием контрастивных функций потерь вида (14), которые позволяют явно оптимизировать модель для выявления семантического сходства предложений, а не просто контекстуальных связей на уровне токенов:

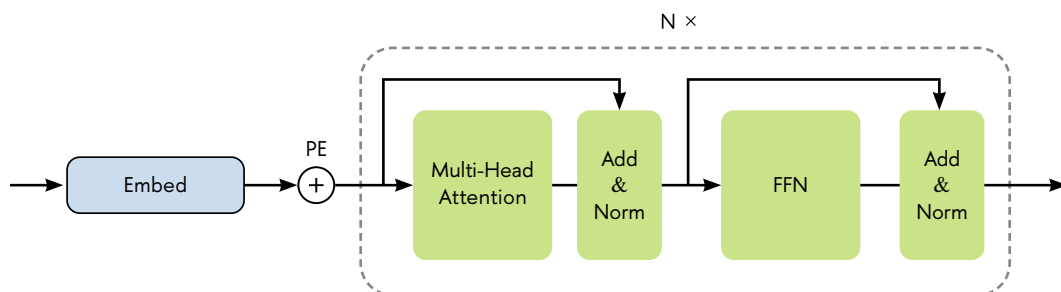


Рис. 1. Архитектура кодировщика.

$$L = (1 - y_k) \|f(x_i) - f(x_j)\|_2^2 + y_k (\max(0, m - \|f(x_i) - f(x_j)\|_2))^2, \quad (14)$$

где x_i и x_j — являются парой объектов;

y_k — бинарная метка, которая принимает значение 0, если объекты являются похожими, и значение 1, если они отличаются;

m — определяет границу разделения между объектами.

Основываясь на вышеизложенном и учитывая поставленную задачу, в данной работе предлагается использовать Sentence-BERT для формирования высококачественных эмбеддингов предложений, которые позволяют эффективно оценивать степень схожести опираясь на семантические связи между названиями товаров, представленных в текстовой форме и, таким образом, получим робастную реализацию компонента ψ_{title} . Используя сгенерированные эмбеддинги, применим в качестве критерия косинусное сходство (15):

$$\text{Sim}(e_i, e_j) = \frac{e_i \cdot e_j}{\|e_i\| \|e_j\|}. \quad (15)$$

Для определения визуальной компоненты модели ψ_{image} необходимо учесть, что на цифровых платформах размещается большое количество разнообразных товаров, которые имеют существенные различия в качестве изображений, ракурсе, освещении, затемнениях и расположении элементов, что в совокупности делает традиционные методы оценки степени схожести недостаточно эффективными. Известно, что такие традиционные сверточные нейронные сети, как ResNet [24] или EfficientNet [25], широко применяются для поиска похожих изображений. Однако они полагаются на локальные свертки и по этой причине имеют фиксированное поле восприятия, что может ограничивать их способность захватывать разрозненные или сильно удаленные друг от друга взаимосвязи между элементами изображения. При мэтчинге товаров важное значение имеют тонкие взаимосвязи между такими элементами, как логотипы, надписи и формы, а CNN часто не удается эффективно справляться с обработкой этих зависимостей. Кроме того, CNN оказываются крайне чувствительны к вышеописанным различиям в изображениях, что также ставит под вопрос эффективность их применения для решения поставленной задачи. Для

преодоления данной проблемы в настоящей работе предлагается использовать архитектуру визуального трансформера (ViT), которая описана в работе [26]. Визуальный трансформер обрабатывает изображения как последовательность фрагментов и использует механизм внимания для фиксации как локальных, так и глобальных взаимосвязей в рамках изображения. В отличие от CNN, которые извлекают признаки иерархическим способом, визуальный трансформер анализирует все изображение одновременно, что позволяет динамически отслеживать наиболее важные области. Эта отличительное свойство делает ViT особенно полезными для оценки сходства предметов одежды, электроники, а также товаров широкого потребления, где различия в текстуре, размещении бренда и различных искажениях могут оказывать существенное влияние на итоговую оценку. Еще одним важным преимуществом визуального трансформера в рамках рассматриваемой задачи является его высокая устойчивость к окклюзиям в изображениях. Например, ViT может правильно идентифицировать модель товара даже если логотип бренда частично скрыт или изображение снято под другим углом, что, как правило, не удается традиционным CNN. Также стоит отметить, что визуальные трансформеры обладают большим потенциалом к интеграции изображения и текстового описания с целью вычисления семантического сходства между ними [27]. Это особенно полезно в сценариях поиска товаров, где различные текстовые характеристики товара (название, описание, марка, модель и так далее) и его изображение могут быть сопоставлены между собой для проверки их соответствия.

Исходя из вышеизложенного в настоящей работе предлагается определить ψ_{image} как многоэтапный конвейер, состоящий из предварительной тренировки, тонкой настройки и выполнения последующей оценки.

На первом этапе проведем начальное обучение на большом объеме данных для формирования обобщенных представлений и выделения базовых признаков. Пусть f_θ задает параметризованный θ визуальный трансформер (ViT) и имеется набор данных, который содержит изображения всех товаров $I = \{x_1, x_2, x_3, \dots, x_n\}$. Применив метод самообучающейся дистилляции (DINO) [28] проведем оптимизацию согласно (16) и для каждого изображения товара сформируем начальный эмбеддинг (17):

$$\theta^* = \arg \min_{\theta} L_{SSL}(f_{\theta}, I). \quad (16)$$

$$e_i = f_{\theta^*}(x_i) \forall x_i \in I. \quad (17)$$

На втором этапе проведем адаптацию визуального трансформера. Для этого используем подмножества изображений товаров, полученные путем разделения I по категориям и ценовым сегментам, выполнив тонкую настройку согласно (18):

$$L = \max \left(\begin{aligned} &\left\| f_{\theta^*}(x_a) - f_{\theta^*}(x_p) \right\|_2^2 - \\ &-\left\| f_{\theta^*}(x_a) - f_{\theta^*}(x_n) \right\|_2^2 + \alpha, 0 \end{aligned} \right), \quad (18)$$

где x_a — является опорным объектом;

x_p — обозначает объект схожий с опорным;

x_n — представляет объект, который является отличным от x_a и x_p .

После завершения тонкой настройки используем (15) для оценки схожести между любой парой векторных представлений e_i и e_j . Описанный конвейер позволит обеспечить высокую эффективность и точность компонента ψ_{image} .

При определении третьего компонента модели $\psi_{metadata}$ необходимо учесть, что характеристики товаров на торговых площадках содержат различные типы данных включая числовые переменные, такие как вес, объем или размер, а также категориальные переменные, такие как комплектация, цвет или состав. Традиционные подходы обычно используют многослойный персептрон (MLP) или ансамблевые методы, которые часто сталкиваются с трудностями при обнаружении комплексного взаимодействия между множеством разнообразных признаков при условии отсутствия предварительного инжиниринга данных. Учитывая характер и специфику предметной области разумно предположить, что определенные признаки, в табличном представлении характеристик товаров, могут взаимодействовать между собой нетривиальным способом. Подобные взаимодействия имеют важное значение в рамках решаемой задачи поскольку даже небольшие различия (например, в составе материалов между «хлопок» и «полихлопок») оказывают существенное влияние на оцениваемую идентичность характеристик. Как было описано ранее, в основе архитектуры трансформера лежит механизм внимания, который позволяет обнаруживать как признаки взаимосвязаны между собой. Контекстно-зависимые эмбединги, создаваемые трансфор-

мером, формируют векторные представления, которые размещают схожие элементы близко друг к другу в пространстве признаков с учетом их взаимосвязей, что позволяет более точно отличать схожие товары на основе имеющихся характеристик. Более того, механизм внимания эффективно выявляет взаимодействия признаков в автоматизированном режиме, что устраняет необходимость участия экспертов предметной области и проведение инжиниринга данных в ручном режиме. Помимо высокой гибкости в отношении признаков, трансформер эффективно справляется с проблемой обработки недостающих или искаженных данных, которая неизбежно возникает на крупных платформах электронной коммерции. Кроме того, некоторые исследования [29, 30], связанные с развитием идеи применения трансформеров для работы с табличными данными, продемонстрировали, что такие архитектуры могут превосходить традиционные подходы.

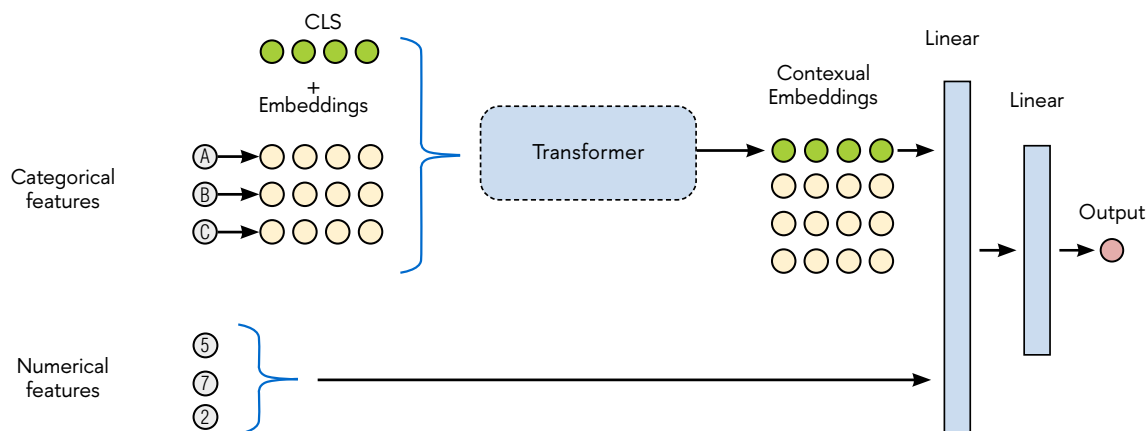
Исходя из вышеизложенного, компонент модели $\psi_{metadata}$ определяется с применением концепции табличного трансформера. Адаптированная для решения поставленной задачи архитектура, предлагаемая в настоящей работе, показана на рисунке 2. Для оценки сходства между любой парой товаров с помощью табличного трансформера будем использовать (15).

Предложенная архитектура табличного трансформера, благодаря использованию контекстно-зависимых эмбедингов для категориальных признаков, специализированной обработки числовых переменных и механизма внимания для выявления сложных взаимодействий, эффективно формирует векторное представление характеристик товара. Это особенно актуально в крупномасштабных и разнородных каталогах, которые являются типичными для цифровых торговых площадок.

После определения всех составляющих итоговую меру сходства зададим с помощью метода взвешенных оценок согласно (19):

$$\begin{aligned} S(p_i, p_j) = & w_1 \text{Sim}(\psi_{title}(p_i), \psi_{title}(p_j)) + \\ & + w_2 \text{Sim}(\psi_{image}(p_i), \psi_{image}(p_j)) + \\ & + w_3 \text{Sim}(\psi_{metadata}(p_i), \psi_{metadata}(p_j)), \sum w_i = 1. \end{aligned} \quad (19)$$

В которой весовые коэффициенты w_1 , w_2 и w_3 могут быть гибко настроены с учетом особенностей и специфики выполняемого мэтинга.

Рис. 2. Архитектура компонента $\psi_{metadata}$

4. Применение модели МММР для оценки схожести товаров на маркетплейсе Wildberries

В рамках проводимого эксперимента было принято решение провести исследование товаров, которые представлены на платформе Wildberries. Основной фокус направлен на товарную категорию «Смартфоны», выбор которой обуславливается рядом значимых факторов. Во-первых, смартфоны представляют собой сложные технические устройства с широким набором характеристик, включая как количественные (например, объем памяти, емкость аккумулятора и так далее), так и категориальные (марка, операционная система и так далее) признаки. Это делает их удобным объектом для тестирования в условиях неоднородности признаков. Во-вторых, ассортимент таких устройств на маркетплейсах достаточно обширен и разнообразен, что позволяет изучить работу модели на данных с высокой степенью вариативности. В-третьих, данная категория является одной из популярных и востребованных на маркетплейсах, что делает результаты исследования практически значимыми. Всё это в совокупности позволяет эффективно проверить возможность определения степени схожести товаров. Таким образом, выбор категории «Смартфоны» позволяет протестировать предложенную модель в условиях реальных данных, обеспечивая достаточную сложность и разнообразие для оценки ее работоспособности. Из этого следует, что успешный опыт применения модели в данной товарной группе может быть перенесен и на другие категории, включая технически сложные товары.

В ходе реализации эксперимента был собран набор данных из 12233 карточек товаров. Каждая карточка содержит информацию о конкретном реализуемом на платформе товаре и включает следующие признаки: X1 — изображение товара, X2 — название товара, X3 — цвет, X4 — модель, X5 — тип SIM карты, X6 — операционная система, X7 — версия операционной системы, X8 — гарантийный срок, X9 — степень пылевлагозащиты, X10 — тип дисплея, X11 — диагональ экрана, X12 — разрешение экрана, X13 — частота обновления экрана, X14 — защитное покрытие экрана, X15 — объем встроенной памяти, X16 — объем оперативной памяти, X17 — основная камера, X18 — фронтальная камера, X19 — особенности объектива, X20 — встроенная вспышка, X21 — емкость аккумулятора, X22 — модель процессора, X23 — количество ядер процессора, X24 — стандарт связи, X25 — беспроводные интерфейсы, X26 — спутниковая навигация, X27 — вид разъема подключения к устройству, X28 — дополнительные опции, X29 — комплектация, X30 — описание товара, X31 — страна производства, X32 — количество SIM карт, X33 — состояние товара, X34 — тактовая частота процессора, X35 — вес товара с упаковкой, X36 — срок эксплуатации.

Фрагмент собранных данных показан в *таблице 1*.

Перед использованием собранных данных для обучения модели была проведена тщательная предварительная обработка, направленная на обеспечение их надлежащего качества, согласованности и пригодности для дальнейшего анализа. Эти работы включали в себя очистку, удаление дубликатов,

Таблица 1.

Фрагмент собранных данных

	Товар 1	Товар 2	Товар 3	Товар 4
Изображение				
Название	Смартфон Apple iPhone 16 Pro Max Золотистый/Desert 512 Гб	Смартфон Apple iPhone 16 Pro Max Белый/White 1 Тб	Смартфон Galaxy S24 Ultra 512 Гб Yellow Желтый	Смартфон Xiaomi 14 12/256 Гб 5G Белый PCT
Цена	168 750	182 267	159 800	77 585
Модель	iPhone 16 Pro Max	iPhone 16 Pro Max	S24	Xiaomi 14
Версия ОС	iOS 18	IOS 18	Android 14	Android 14
Гарантийный срок	1 год	1 год	1 год	1 год
Тип дисплея	Super Retina XDR OLED ProMotion	Super Retina XDR OLED ProMotion	Dynamic AMOLED 2X	LTPO AMOLED
Диагональ экрана	6.9"	6.9"	6.8"	6.36"
Разрешение экрана	2868x1320	2868x1320	3120x1440	2670x1200
Объем встроенной памяти	512 Гб	1 Тб	512 Гб	256 Гб
Объем оперативной памяти	8 Гб	8 Гб	12 Гб	12 Гб
Беспроводные интерфейсы	Wi-Fi 7; NFC; Bluetooth 5.3	Wi-Fi 7; NFC; Bluetooth 5.3	Wi-Fi; NFC; Bluetooth	Wi-Fi; ИК-Порт; NFC; Bluetooth
...	...	...	...	...

обработку пропущенных значений, исправление несоответствий, нормализацию числовых данных, а также токенизацию текстовой информации. Дополнительно для минимизации влияния аномальных значений была проведена работа по обнаружению и удалению выбросов.

После обучения модели были получены оценки схожести для товаров, показанных в *таблице 1*, результаты которых приведены в *таблице 2*. Эффективность предложенного метода была также

оценена путём сопоставления с наивным базовым подходом, реализованного на основе сравнения товаров с использованием коэффициента Жаккара без предварительной токенизации.

Экспериментальная часть исследования подтвердила эффективность разработанного подхода. В частности, данные *таблицы 2* демонстрируют пример оценок сходства, рассчитанных моделью МММР для четырёх товарных позиций (Товар 1, Товар 2, Товар 3 и Товар 4). Наиболее высокое зна-

Таблица 2.

Оценка схожести товаров

		Товар 1	Товар 2	Товар 3	Товар 4
Товар 1	MMMP	1	0,98657	0,877564	0,544109
	Jaccard	1	0,308	0,11	0,077
Товар 2	MMMP	0,98657	1	0,883621	0,549463
	Jaccard	0,308	1	0,10	0,11
Товар 3	MMMP	0,877564	0,883621	1	0,670598
	Jaccard	0,11	0,10	1	0,13
Товар 4	MMMP	0,544109	0,549463	0,670598	1
	Jaccard	0,077	0,11	0,13	1

чение сходства (0,98657) наблюдается между Товаром 1 и Товаром 2, что свидетельствует о правильной идентификации практически идентичных позиций при наличии различных описаний с использованием различных модальностей данных. При этом менее похожие позиции (например, Товар 1 и Товар 4) получили намного более низкий коэффициент сходства (0,544109), что говорит о корректном разграничении моделью групп товаров с разным набором характеристик и описаний. Аналогичные результаты (0,549463 между Товаром 2 и Товаром 4, а также 0,670598 между Товаром 3 и Товаром 4) показывают, что модель способна отличить даже незначительные различия в характеристиках товаров и при этом способна выявлять интересные паттерны в их описании. Следовательно, предложенный подход надёжно справляется с выявлением как очевидных, так и более тонких совпадений и при этом не позволяет объединить записи о товарах в одну группу, если товары в действительности различаются.

Полученные результаты подтверждают, что мультимодальная модель на базе трансформера способна эффективно решать задачу мэтчинга в условиях большого разнообразия товарных позиций при разнородности источников данных. Корректная идентификация записей о товарных позициях на маркетплейсах сокращает риски дублирования и ошибок в аналитике, положительно сказываясь на всех этапах экономической деятельности в рамках экосистемы цифровых платформ, а также позволяет продавцам и покупателям более точно ориентироваться в ассортименте, упрощая

процесс принятия решений. Кроме того, полученные результаты свидетельствуют о корректной работе модели в рамках поставленной задачи, открывая возможность ее применения для интеллектуального мэтчинга товарных предложений на маркетплейсах.

Заключение

В ходе проведённого исследования была предложена и решена задача интеллектуального мэтчинга товаров на маркетплейсах, требующая комплексного анализа и обработки мультимодальных данных и применения современных инструментальных методов в рамках построения экономико-математической модели (MMMP). В первом разделе статьи был раскрыт потенциал применения мэтчинга товаров на маркетплейсах, как типичных представителей цифровых платформ, и показана взаимосвязь между возможностями интеллектуального мэтчинга и ключевыми экономическими показателями деятельности в рамках рынка электронной коммерции. Проведенный во втором разделе анализ существующих подходов к entity resolution (ER) и анализу мультимодальных данных позволил выявить ключевую проблему сопоставления записей о товарах на маркетплейсах смысл которой заключается в том, что товары описываются различными продавцами с разным уровнем детализации и в разных форматах (текст, изображения, табличные характеристики и т.д.). В третьем разделе описан детальный процесс разработки мультимодальной модели мэтчинга (MMMP), ядром которой явля-

ются модули на базе трансформеров для обработки текстовых, визуальных и табличных данных. Описанная архитектура учитывает гибкость механизма внимания и обладает способностью к самообучению при расширении ассортимента товаров на маркетплейсах и появлению новых неструктурированных описаний с учетом визуальных компонент данных. МММР позволяет эффективно интегрировать многообразие входных модальностей данных, а также учитывать сложные контекстные взаимосвязи, важные для корректного определения сходства или различия товарных позиций с учетом ценовых характеристик.

Наконец, в четвертом разделе было продемонстрировано применение предложенной модели для оценки схожести товарных позиций на маркетплейсе Wildberries, где полученные результаты подтвердили высокую точность и стабильность работы МММР.

Таким образом, основные выводы по статье можно представить в аксиоматической форме:

- ♦ обоснована высокая релевантность мультимодального анализа для задач мэтчинга в условиях растущего многообразия товарных позиций на маркетплейсах;
- ♦ определена перспективность применения архитектур трансформера для комплексной обработки текстовых, визуальных и табличных признаков;
- ♦ выявлена необходимость дальнейшей адаптации подобных моделей к изменяющимся условиям рынка;

♦ разработана модель МММР, которая может использоваться как в научных исследованиях по интеллектуальной идентификации записей о товарах на маркетплейсах, так и в практических целях участниками рынка электронной коммерции, где точность сопоставления записей о товарах является критически важной, например, для ценообразования, а значит и для формирования равновесной модели спроса и предложения.

Результаты работы доказывают, что грамотно спроектированная архитектура модели глубокого обучения, интегрирующая в себе несколько модальностей данных, даёт ощутимое преимущество по сравнению с более простыми и узкоспециализированными существующими решениями. Кроме того, применение подобных моделей содействует повышению прозрачности и эффективности аналитических инструментов, а это, в конечном итоге, будет укреплять доверие продавцов и, что более важно доверие покупателей к определенному маркетплейсу в целом.

Дальнейшие пути исследования в рамках выбранной тематики могут заключаться в расширении набора обрабатываемых модальностей (например, анализ видеозаписей или аудиоинформации) и в создании самообучающихся механизмов, позволяющих автоматически перенастраивать модель при изменении структуры данных. Также вектор дальнейшего развития модели МММР предполагает интеграцию методов активного обучения (в реальном режиме времени), поскольку это позволит быстрее накапливать и обрабатывать релевантные примеры для уточнения критериев мэтчинга, в том числе и на маркетплейсах. ■

Литература

1. Fletcher A., Ormosi P.L., Savani R. Recommender systems and supplier competition on platforms // *Journal of Competition Law & Economics*. 2023. Vol. 19. No. 3. P. 397–426. <https://doi.org/10.1093/joclec/nhad009>
2. Hussien F.T.A., Rahma A.M.S., Abdulwahab H.B. An e-commerce recommendation system based on dynamic analysis of customer behavior // *Sustainability*. 2021. Vol. 13. No. 19. Article 10786. <https://doi.org/10.3390/su131910786>
3. Studying product competition using representation learning / F. Chen [et al.] // *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '20)*. 2020. P. 1261–1268. <https://doi.org/10.1145/3397271.3401041>
4. Hu S., Wei M.M., Cui S. The role of product and market information in an online marketplace // *Production and Operations Management*. 2023. Vol. 32. No. 10. P. 3100–3118. <https://doi.org/10.1111/poms.14025>
5. Detecting online counterfeit-goods seller using connection discovery / M. Cheung [et al.] // *ACM Transactions on Multimedia Computing, Communications, and Applications*. 2019. Vol. 15. No. 2. Article 35. <https://doi.org/10.1145/3311785>
6. Sun J., Zhang X., Zhu Q. Counterfeiters in online marketplaces: Stealing your sales or sharing your costs // *Journal of Retailing*. 2020. Vol. 96. No. 2. P. 189–202. <https://doi.org/10.1016/j.jretai.2019.07.002>

7. Köpcke H., Thor A., Rahm E. Evaluation of entity resolution approaches on real-world match problems // *Proceedings of the VLDB Endowment*. 2010. Vol. 3. Nos. 1–2. P. 484–493. <https://doi.org/10.14778/1920841.1920904>
8. Cohen W.W., Ravikumar P., Fienberg S.E. A comparison of string distance metrics for name-matching tasks // *Proceedings of Workshop on Information Integration (IJCAI-03)*. 2003. P. 73–78.
9. Synthesizing entity matching rules by examples / R. Singh [et al.] // *Proceedings of the VLDB Endowment*. 2017. Vol. 11. No. 2. P. 189–202. <https://doi.org/10.14778/3149193.3149199>
10. Entity matching: How similar is similar / J. Wang [et al.] // *Proceedings of the VLDB Endowment*. 2011. Vol. 4. No. 10. P. 622–633. <https://doi.org/10.14778/2021017.2021020>
11. Angermann H. TaxoMulti: Rule-based expert system to customize product taxonomies for multi-channel e-commerce // *SN Computer Science*. 2022. Vol. 3. Article 177. <https://doi.org/10.1007/s42979-022-01070-8>
12. Hybrid ecommerce recommendation model incorporating product taxonomy and folksonomy / M. Mao [et al.] // *Knowledge-Based Systems*. 2021. Vol. 214. Article 106720. <https://doi.org/10.1016/j.knosys.2020.106720>
13. Aanen S.S., Vandic D., Frasincar F. Automated product taxonomy mapping in an e-commerce environment // *Expert Systems with Applications*. 2015. Vol. 42. No. 3. P. 1298–1313. <https://doi.org/10.1016/j.eswa.2014.09.032>
14. A machine learning approach for product matching and categorization: Use case: Enriching product ads with semantic structured data / P. Ristoski [et al.] // *Semantic Web*. 2018. Vol. 9. No. 4. P. 707–728. <https://doi.org/10.3233/SW-180300>
15. Shah K., Kopru S., Ruvini J.D. Neural network based extreme classification and similarity models for product matching // *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, New Orleans – Louisiana*. 2018. Vol. 3. P. 8–15. Association for Computational Linguistics. <https://doi.org/10.18653/v1/N18-3002>
16. Attention is all you need / A. Vaswani [et al.] // *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*, Long Beach, CA, USA. 2017. P. 6000–6010.
17. Zhang H., Shafiq M.O. Survey of transformers and towards ensemble learning using transformers for natural language processing // *Journal of Big Data*. 2024. Vol. 11. Article 25. <https://doi.org/10.1186/s40537-023-00842-0>
18. Efficient estimation of word representations in vector space / T. Mikolov [et al.] // *arXiv:1301.3781*. 2013. <https://doi.org/10.48550/arXiv.1301.3781>
19. Pennington J., Socher R., Manning C.D. Glove: Global vectors for word representation // *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar. 2014. P. 1532–1543.
20. Deep residual learning for image recognition / He K. [et al.] // *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA. 2016. P. 770–778. <https://doi.org/10.1109/CVPR.2016.90>
21. Ba J.L., Kiros J.R., Hinton G.E. Layer normalization. *arXiv:1607.06450*. 2016. <https://doi.org/10.48550/arXiv.1607.06450>
22. Bert: Pre-training of deep bidirectional transformers for language understanding / Devlin J. [et al.] // *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, Minnesota*. 2019. Vol. 1. P. 4171–4186.
23. Reimers N., Gurevych I. Sentence-BERT: Sentence embeddings using Siamese BERT-networks // *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, China. Association for Computational Linguistics. 2019. P. 3982–3992. <https://doi.org/10.18653/v1/D19-1410>
24. Wu Z., Shen C., van den Hengel A. Wider or deeper: Revisiting the ResNet model for visual recognition // *Pattern Recognition*. 2019. Vol. 90. P. 119–133.
25. Tan M., Le Q. EfficientNet: Rethinking model scaling for convolutional neural networks // *Proceedings of the 36th International Conference on Machine Learning*. 2019. Vol. 97. P. 6105–6114.
26. An image is worth 16x16 words: Transformers for image recognition at scale / A. Dosovitskiy [et al.] // *arXiv:2010.11929*. 2021. <https://doi.org/10.48550/arXiv.2010.11929>
27. Learning transferable visual models from natural language supervision / A. Radford [et al.] // *Proceedings of the 38th International Conference on Machine Learning*. 2021. Vol. 139. P. 8748–8763.
28. Emerging properties in self-supervised vision transformers / M. Caron [et al.] // *arXiv:2104.14294*. 2021. <https://doi.org/10.48550/arXiv.2104.14294>
29. TabTransformer: Tabular data modeling using contextual embeddings / X. Huang [et al.] // *arXiv:2012.06678*. 2020. <https://doi.org/10.48550/arXiv.2012.06678>
30. Revisiting deep learning models for tabular data / Y. Gorishniy [et al.] // *Proceedings of the 35th International Conference on Neural Information Processing Systems (NIPS'21)*. 2021. Article 1447. P. 18932–18943.

Об авторах

Варнухов Артем Юрьевич

ассистент, кафедра бизнес-информатики, Уральский государственный экономический университет, Россия, 620144, г. Екатеринбург, ул. 8 Марта, д. 62;
E-mail: varnuhov_ayu@usue.ru

Назаров Дмитрий Михайлович

доктор экономических наук;
заведующий кафедрой, кафедра бизнес-информатики, Уральский государственный экономический университет, Россия, 620144, г. Екатеринбург, ул. 8 Марта, д. 62;
E-mail: slup2005@mail.ru
ORCID: 0000-0002-5847-9718

Product matching in digital marketplaces: Multimodal model based on the transformer architecture

Artem Yu. Varnukhov

E-mail: varnuhov_ayu@usue.ru

Dmitry M. Nazarov

E-mail: slup2005@mail.ru

Ural State University of Economics, Yekaterinburg, Russia

Abstract

In this paper we analyze the problem of intelligent product matching in digital marketplaces for which one requires evaluation of similarity of various records that describe products but may differ in format, content or volume of multimodal data. The subject area of this scientific research represents an intersection of entity resolution (ER) problem solving methods: record matching and multimodal data analysis. It is of extreme relevance in a fast-growing platform economy with the e-commerce market expanding exponentially. The main purpose of this research is to develop and test an intelligent multimodal model based on transformer architecture to improve the accuracy and robustness of product matching in digital marketplaces. The authors developed a model integrating textual, visual and tabular attributes which enables us to identify similar products, find competitive offers, detect duplicates and perform product clustering and segmentation in a more effective manner. The proposed approach is based on the self-attention mechanism which enables contextual-semantic relations modeling of various-nature data. In

order to extract the vector representation of text descriptions, language models are applied, in particular the Sentence-BERT architecture; for the graphical component Vision Transformer is used; and tabular data are processed using specialized learning mechanisms based on TabTransformer structured data. The experiment we carried out demonstrated that the developed multimodal model efficiently solves the task of product matching in digital marketplaces in an environment of significant variability of product items and data heterogeneity. Additionally, the results suggest that the model can be adapted successfully for application in other product categories. The results obtained confirm the efficiency and expediency to apply the multimodal approach for digital marketplace product matching implementation. This allows the e-commerce market participants to significantly improve the quality of inventory management, increase pricing efficiency and strengthen their competitive advantages.

Keywords: digital marketplace, contextual-semantic identification, competitive offers search, product matching, machine learning, deep learning, transformer architecture, data mining

Citation: 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

References

1. Fletcher A., Ormosi P. L., Savani R. (2023) Recommender systems and supplier competition on platforms. *Journal of Competition Law & Economics*, vol. 19, no. 3, pp. 397–426. <https://doi.org/10.1093/joclec/nhad009>
2. Hussien F.T.A., Rahma A.M.S., Abdulwahab H.B. (2021) An e-commerce recommendation system based on dynamic analysis of customer behavior. *Sustainability*, vol. 13, no. 19, article 10786. <https://doi.org/10.3390/su131910786>
3. Chen F., Liu X., Proserpio D., et al. (2020) Studying product competition using representation learning. *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '20)*, pp. 1261–1268. <https://doi.org/10.1145/3397271.3401041>
4. Hu S., Wei M.M., Cui S. (2023) The role of product and market information in an online marketplace. *Production and Operations Management*, vol. 32, no. 10, pp. 3100–3118. <https://doi.org/10.1111/poms.14025>
5. Cheung M., She J., Sun W., Zhou J. (2019) Detecting online counterfeit-goods seller using connection discovery. *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 15, no. 2, article 35. <https://doi.org/10.1145/3311785>
6. Sun J., Zhang X., Zhu Q. (2020) Counterfeiters in online marketplaces: Stealing your sales or sharing your costs. *Journal of Retailing*, vol. 96, no. 2, pp. 189–202. <https://doi.org/10.1016/j.jretai.2019.07.002>
7. Köpcke H., Thor A., Rahm E. (2010) Evaluation of entity resolution approaches on real-world match problems. *Proceedings of the VLDB Endowment*, vol. 3, nos. 1–2, pp. 484–493. <https://doi.org/10.14778/1920841.1920904>
8. Cohen W.W., Ravikumar P., Fienberg S.E. (2003) A Comparison of string distance metrics for name-matching tasks. *Proceedings of Workshop on Information Integration (IJCAI-03)*, pp. 73–78.
9. Singh R., Meduri V.V., Elmagarmid A., et. al. (2017) Synthesizing entity matching rules by examples. *Proceedings of the VLDB Endowment*, vol. 11, no. 2, pp. 189–202. <https://doi.org/10.14778/3149193.3149199>
10. Wang J., Li G., Yu J.X., Feng J. (2011) Entity matching: How similar is similar. *Proceedings of the VLDB Endowment*, vol. 4, no. 10, pp. 622–633. <https://doi.org/10.14778/2021017.2021020>
11. Angermann H. (2022) TaxoMulti: Rule-based expert system to customize product taxonomies for multi-channel e-commerce. *SN Computer Science*, vol. 3, article 177. <https://doi.org/10.1007/s42979-022-01070-8>
12. Mao M., Chen S., Zhang F., et. al. (2021) Hybrid ecommerce recommendation model incorporating product taxonomy and folksonomy. *Knowledge-Based Systems*, vol. 214, article 106720. <https://doi.org/10.1016/j.knosys.2020.106720>
13. Aanan S. S., Vandic D., Frasincar F. (2015) Automated product taxonomy mapping in an e-commerce environment. *Expert Systems with Applications*, vol. 42, no. 3, pp. 1298–1313. <https://doi.org/10.1016/j.eswa.2014.09.032>
14. Ristoski P., Petrovski P., Mika P., Paulheim H. (2018) A machine learning approach for product matching and categorization: Use case: Enriching product ads with semantic structured data. *Semantic Web*, vol. 9, no. 5, pp. 707–728. <https://doi.org/10.3233/SW-180300>

15. Shah K., Kopru S., Ruvini J. D. (2018) Neural network based extreme classification and similarity models for product matching. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, New Orleans – Louisiana*, vol. 3, pp. 8–15. Association for Computational Linguistics. <https://doi.org/10.18653/v1/N18-3002>
16. Vaswani A., Shazeer N., Parmar N., et. al. (2017) Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), Long Beach, CA, USA*, pp. 6000–6010. <https://dl.acm.org/doi/pdf/10.5555/3295222.3295349>
17. Zhang H., Shafiq M.O. (2024) Survey of transformers and towards ensemble learning using transformers for natural language processing. *Journal of Big Data*, vol. 11, article 25. <https://doi.org/10.1186/s40537-023-00842-0>
18. Mikolov T., Chen K., Corrado G., Dean J. (2013) Efficient estimation of word representations in vector space. *arXiv:1301.3781*. <https://doi.org/10.48550/arXiv.1301.3781>
19. Pennington J., Socher R., Manning C. D. (2014) GloVe: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar*, pp. 1532–1543.
20. He K., Zhang X., Ren S., Sun J. (2016) Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA*, pp. 770–778. <https://doi.org/10.1109/CVPR.2016.90>
21. Ba J.L., Kiros J.R., Hinton G.E. (2016) Layer normalization. *arXiv:1607.06450*. <https://doi.org/10.48550/arXiv.1607.06450>
22. Devlin J., Chang M. W., Lee K., Toutanova K. (2019) Bert: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, Minnesota*, vol. 1, pp. 4171–4186. Association for Computational Linguistics. <https://doi.org/10.18653/v1/N19-1423>
23. Reimers N., Gurevych I. (2019) Sentence-BERT: Sentence embeddings using Siamese BERT-networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China*, pp. 3982–3992. Association for Computational Linguistics. <https://doi.org/10.18653/v1/D19-1410>
24. Wu Z., Shen C., van den Hengel A. (2019) Wider or deeper: Revisiting the ResNet model for visual recognition. *Pattern Recognition*, vol. 90, pp. 119–133. <https://doi.org/10.1016/j.patcog.2019.01.006>
25. Tan M., Le Q. (2019) EfficientNet: Rethinking model scaling for convolutional neural networks. *Proceedings of the 36th International Conference on Machine Learning*, vol. 97, pp. 6105–6114.
26. Dosovitskiy A., Beyer L., Kolesnikov A., et al. (2021) An image is worth 16x16 words: transformers for image recognition at scale. *arXiv:2010.11929*. <https://doi.org/10.48550/arXiv.2010.11929>
27. Radford A., Kim J. W., Hallacy C., et. al. (2021) Learning transferable visual models from natural language supervision. *Proceedings of the 38th International Conference on Machine Learning*, vol. 139, pp. 8748–8763.
28. Caron M., Touvron H., Misra I., et. al. (2021) Emerging properties in self-supervised vision transformers. *arXiv:2104.14294*. <https://doi.org/10.48550/arXiv.2104.14294>
29. Huang X., Khetan A., Cvitkovic M., et. al. (2020) TabTransformer: Tabular data modeling using contextual embeddings. *arXiv:2012.06678*. <https://doi.org/10.48550/arXiv.2012.06678>
30. Gorishniy Y., Rubachev I., Khrulkov V., et. al. (2021) Revisiting deep learning models for tabular data. *Proceedings of the 35th International Conference on Neural Information Processing Systems (NIPS'21)*, article 1447, pp. 18932–18943.

About the authors

Artem Yu. Varnukhov

Assistant, Department of Business Informatics, Ural State University of Economics, 62, 8 Marta St., Yekaterinburg 620144, Russia;
E-mail: varnuhov_ayu@usue.ru

Dmitry M. Nazarov

Doctor of Sciences (Economics);
Head of Department, Department of Business Informatics, Ural State University of Economics, 62, 8 Marta St., Yekaterinburg 620144, Russia;
E-mail: slup2005@mail.ru
ORCID: 0000-0002-5847-9718