

Prod2Query: решение проблемы холодного старта продаж для электронной коммерции с помощью генеративного языкового моделирования

Ф.В. Краснов

Аннотация— Крупные торговые интернет-площадки представляют тысячи новых товаров ежедневно. Чтобы приобрести новый товар пользователь должен найти его через систему поиска, а системы поиска сейчас в своем большинстве используют при индексации товаров поведенческие сигналы, такие как, покупка, клик, просмотр товара, которых еще нет для новых товаров. В этом и состоит проблема «холодного» старта продаж. С появлением генеративных языковых моделей стало возможным обучить генеративную языковую модель на поведении пользователей для того, чтобы сгенерировать поведенческие поисковые запросы для новых товаров. Таким образом, получается набор синтетических поведенческих данных для новых товаров, который можно использовать для обучения системы поиска. Основная исследовательская задача статьи состоит в том, чтобы проверить насколько улучшаются автономные показатели системы поиска, обученной на таком синтетическом наборе данных созданном для новых товаров. В основе Prod2Query лежит архитектура модели Encoder-Decoder на трансформерах. В результате обработки новых товаров с помощью Prod2Query получен показатель для модели извлечения новых товаров $mAP@12 = 77.2\%$, что находится на уровне SOTA моделей. Такой результат означает, что проблему «холодного» старта можно эффективно решать на основе сигналов от продавцов при заведении новых товаров путем формирования синтетических поисковых запросов для обучения моделей поиска.

Ключевые слова — LLM, холодный старт продаж, продуктовый поиск, системы извлечения товаров.

I. ВВЕДЕНИЕ

Проблема холодного старта продаж активно изучается в рекомендательных системах [1, 2, 3], но исследований, посвященных системам поиска товаров, довольно мало. Причина в том, что при лексическом поиске появление нового товара не означает появление нового токена в словаре. Другими словами, новый продавец товара с названием «чехол для huawei» не создает проблему холодного старта для системы поиска, если товары с названием «чехол для huawei» ранее уже покупались у других продавцов. Поэтому исследователи уделяют больше внимание ранжированию товаров от новых, «холодных» продавцов. Работы [4, 5] описывают, как искусственно создать отзывы клиентов

для новых товаров, чтобы эмулировать сигналы вовлеченности клиентов, влияющие на появление товара в поисковой выдаче. В исследовании [6] приведен опыт борьбы с холодным стартом в поисковой системе Airbnb. Предложенные подходы сосредоточены на введении признаков для новых товаров, которые влияют на поисковую выдачу в течение определенного времени. В [7] применены стратегии трансфертного обучения для моделей ранжирования поиска в новых категориях товаров, где сигналы от покупателей еще не доступны. Приведенные исследования ориентированы на улучшение точности поисковой выдачи для новых товаров. С другой стороны, есть ряд экспериментов по включению сигналов пользователей в подсистему поиска, предназначенную для извлечения информации о товарах. Например, в исследовании [8] при извлечении товаров по запросу использованы сигналы от просмотров и покупок по цифровым идентификаторам товаров. При таком подходе сохраняется проблема холодного старта, так как цифровой идентификатор товара будет уникальным, и тот факт, что название товара уже существует в индексе, будет иметь ограниченное влияние на выдачу товаров-кандидатов.

Нейросетевые модели, особенно основанные на больших языковых моделях, повышают эффективность поиска [9]. Исследователи разделяют подходы, направленные на повторное ранжирование извлеченных товаров для повышения точности [10], и подходы, нацеленные на улучшение самого извлечения товаров [11]. Например, в работе [12] улучшен подход извлечения товаров с помощью аугментации. Этот подход обучает модель «последовательность-последовательность» для предсказания запросов, которые могут быть релевантны конкретному документу (работа посвящена информационному поиску по документам). При индексировании эта модель используется для расширения документа путем создания коллекции запросов и добавления их в документ. Расширение документа приводит к дополнительным вычислительным затратам во время индексирования [13], но авторы [12] отмечают рост показателей точности поиска даже при использовании лексических моделей, таких как BM25 [14]. Многочисленные работы показали, что аугментирование поисковых запросов позволяет

повысить эффективность последующих этапов ранжирования [15-18].

Однако модели «последовательность-последовательность» склонны генерировать контент, не отражающий входной текст. Этот недостаток известен в научной литературе как «галлюцинация» [19]. «Галлюцинации» модели негативно сказываются на качестве машинных переводов и «вопрос-ответных» систем, когда сгенерированный результат воспринимает человек. Но в процессе продуктового поиска сгенерированные поисковые запросы использует только модель при индексировании товаров, потому влияние «галлюцинаций» не столь заметно.

Задача данного исследования — измерить эффективность применения генеративных языковых моделей по методике Prod2Query для решения проблемы холодного старта. Для этого сформированы данные о поведенческих сигналах пользователей, обучены модели по методике Prod2Query, проведен эксперимент на базе 800 тыс. поисковых запросов, на основе автономных показателей системы извлечения информации о товарах оценена эффективность предложенного решения.

II. МЕТОДИКА

Методика Prod2Query включает фазы обучения, генерации и оценки. Для обучения Prod2Query необходимо настроить сбор данных, отражающих поведение пользователей при совершении покупки. Процесс поиска товара сопровождается введением поисковых запросов, просмотром предложенных товаров и добавлением понравившихся в корзину для последующей покупки. В отношении сигнала к шуму наилучшим выражением намерения в поисковом запросе являются поисковые запросы, приводящие к покупке товара.

Для генерации поисковых запросов на вход обученной модели Prod2Query подаются текстовые представления товаров. Модель в рамках Prod2Query генерирует набор из нескольких поисковых запросов, в разной степени соответствующих текстовому представлению товара. Так, часть принципиально важных для данного товара характеристик может не войти в сгенерированный поисковый запрос. Или поисковый запрос может оказаться слишком длинным и специфичным, что также сделает поиск по данному запросу не эффективным. На практике часто используют лексический шаблон для создания поискового запроса на основании текстового представления товара. Например, поисковый запрос по шаблону: «бренд» + «название товара» + «характеристики» будет выглядеть так: «Платочная мануфактура Россия Палантин шарф легкий однотонный 170x60». Шаблоны состоят из правил, и

при широком товарном ассортименте правил становится настолько много, что управление ими превращается в отдельную сложную задачу. Поэтому рассмотрена возможность моделирования этого процесса с помощью языковой трансформации.

В основе Prod2Query лежит генеративная языковая модель. Для практической реализации оценки обученности генеративных систем используют метрики на основе сопоставления слов между сгенерированным текстом и истинным. Применяют следующие показатели совпадения текста: BLEU [20], NIST [21], ROUGE [22], TER [23], METEOR [24]. Один из их основных принципов — штраф за краткость. Лучше, если сгенерированный ответ состоит из большего количества токенов, чем истинный. В случае Prod2Query истинный текст поискового запроса не обязательно является оптимальным для данного товара, так как пользователь может ввести большое количество избыточных слов. Таким образом, краткость, наоборот, может поощряться при оценке эффективности модели. Для управления длиной генерируемого поискового запроса по методике Prod2Query используется параметр ширины лучевого поиска $beam_{width}$ [25].

Наиболее репрезентативный показатель для данного исследования — TER. Он определяется как минимальное количество правок, необходимое для изменения сгенерированного запроса при точном совпадении с оригинальным поисковым запросом, нормализованное по средней длине сгенерированных поисковых запросов.

На рисунке 1 приведена схема обучения и генерации модели по методике Prod2Query.

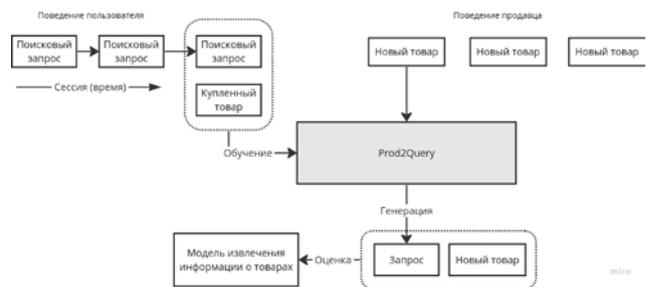


Рис.1. Схема обучения и генерации модели по методике Prod2Query

В поиске по товарам основными автономными показателями являются пороговая полнота и точность. Для оценки модели выбран показатель пороговой точности $mAP@k$, схема его использования приведена на рисунке 2.

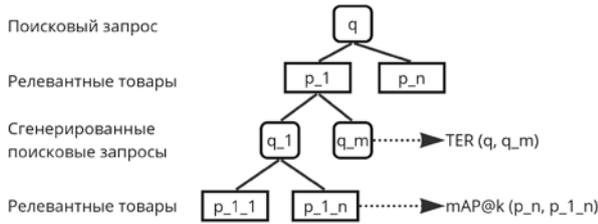


Рис. 2. Схема использования показателей mAP@k и TER в Prod2Query

В исследованиях [26, 27] показатель пороговой точности mAP@k применен для автономной оценки систем извлечения информации о товарах по поисковому запросу. В настоящем исследовании mAP@k позволяет оценить набор поисковых запросов для существующей системы извлечения товаров, поэтому для mAP@k использованы формулы (1) и (2).

$$mAP_m@k = \frac{1}{k} \sum_{i=1}^k P_m@i \quad (2) P_m@k$$

$$= \frac{1}{|Q|} \sum_{q_i \in Q} \frac{|M_m(p_{q_i}^g, p_{q_i}^r@k)|}{k} \quad (1)$$

$|Q|$ – количество рассматриваемых поисковых запросов,

k – порог отсечки поисковой выдачи,

q_i – поисковый запрос $q_i \in Q$,

$p_{q_i}^g$ – все товары соответствующие поисковому запросу q_i ,

$p_{q_i}^r$ – поисковая выдача, все товары найденные по поисковому запросу q_i ,

$|M_m(\cdot, \cdot)|$ – количество соответствующих друг другу товаров, определенных по функции соответствия M для модальности m .

Показатель mAP@k является мультимодальным. Выбрана модальность m , соответствующая цифровым идентификаторам товаров. Функция $M_m(\cdot, \cdot)$ определена как $M_m(p_{q_i}, p_{q_j}) = p_{q_i} \cap p_{q_j}$, где p_{q_i} и p_{q_j} – цифровые идентификаторы товаров, полученные по запросам q_i и q_j соответственно.

В промышленной эксплуатации модель, обученная по методике Prod2Query, служит моделью-учителем [28], а модель извлечения информации о товарах на основе поведения пользователя – моделью-учеником.



Рис. 3. Схема «учитель-ученик» для использования Prod2Query

Моделью-учеником может быть, например, DSSM [29] или Dual Encoder [30], относящиеся к классу моделей Embedding-based Retrieval (EBR) [31]. Схема «учитель-ученик» в рамках Prod2Query представлена на

рисунке 3.

III. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

В качестве множества запросов Q в эксперименте выбрано 800 тыс. наиболее частотных запросов. К примеру, «футболка женская с длинным рукавом хлопок большой размер». По каждому поисковому запросу извлечено 12 товаров-кандидатов. Для каждого товара-кандидата с помощью модели Prod2Query сгенерированы поисковые запросы с различными значениями длины генерируемого запроса $beam_{width} \in [3,5,7,9]$.

Для обучения модели по методике Prod2Query в эксперименте использован программный каркас [32] со следующими параметрами (таблица 1):

Таблица 1 — Значения параметров модели, обученной по методике Prod2Query

Параметр	Значение
layers	4
hidden_size	128
word_vec_size	128
max_grad_norm	0
optim	adam
encoder_type	transformer
decoder_type	transformer
position_encoding	true
dropout	0.1
attention_dropout	0.1
shared_vocab	true

В качестве обучающего набора данных выбраны поисковые запросы, приведшие к покупке, всего 52 млн пар «запрос – товар». В качестве текстовых представлений товара использованы «Товарная категория», «Название», «Бренд», «Характеристики товара». Примеры сгенерированных запросов приведены в таблице 2.

Таблица 2. Примеры сгенерированных по методике Prod2Query запросов

Поисков	Сгенерированные запросы		
ый запрос	$beam_{width} = 5$	$beam_{width} = 7$	$beam_{width} = 9$
штормы	тюль	тюль вуаль	тюль в гостиную
для	200x240	200x240	200x240 готовая
гостиной			на ленте
200 270			
готовые			

две шт

костюм брючный женский нарядный стильный большой размер	брючный костюм женский праздничный	брючный костюм женский праздничный	брючный костюм с пиджаком и пиджаком
ортопеди сабо ческая обувь женская медицинс кая с открыты м носом	кроксы медицинские женские	кроксы медицинские женские с закрытым носом	
папка труда с ручками для девочки	папка для труда	папка для труда a4	папка для труда a4 на молнии с ручками
рулетка для собак 5 м flexi лента	поводок flexi рулетка	поводок flexi рулетка для собак	

Генеративная модель берет параметр длины запроса в качестве ориентира, но запросы получаются различной длины. Для проверки этого явления построена зависимость количества токенов в сгенерированном запросе, приведенная на рисунке 4.

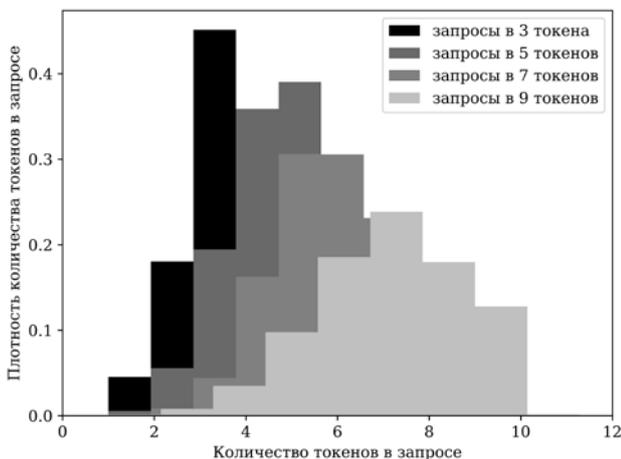


Рис. 4. Зависимость количества токенов в сгенеренном запросе от параметра $beam_{width}$

Качество сгенерированных запросов оценено по показателю TER с помощью программного каркаса SacreBLEU [33]. На рисунке 5 построено распределение значений показателя TER в зависимости от параметра $beam_{width}$.

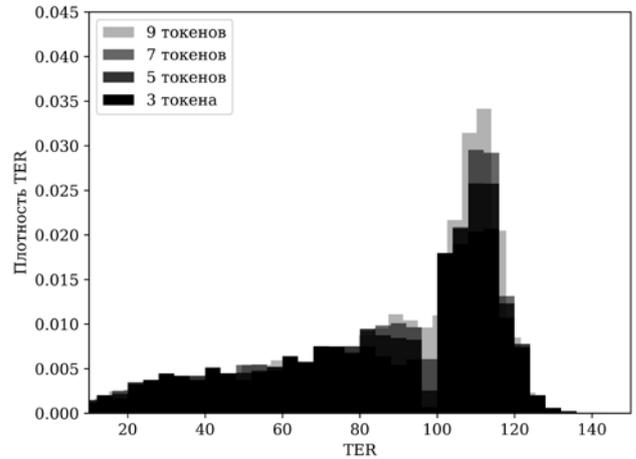


Рис. 5. Распределение значений показателя TER в зависимости от параметра $beam_{width}$

Форма распределения TER на рисунке 5 не зависит от параметра $beam_{width}$. Это означает, что сгенерированные запросы не различимы по показателю TER.

По каждому сгенерированному запросу извлечено n товаров-кандидатов согласно схеме на рисунке 2. На основании полученных товаров-кандидатов проанализирован показатель точности $mAP@12$ (таблица 3).

Таблица 3 — Показатель точности $mAP@12$ для различных значений параметра $beam_{width}$

$beam_{width}$	$mAP@12$
3	76.0 %
5	77.2 %
7	67.2 %
9	60.6%

Оценена точность получения запросов из названий и характеристик товаров на основании шаблона: $mAP@12 = 67\%$. Для сравнения в исследовании [30] значение mAP составляет 74.5%. Полученные результаты показывают, что при $beam_{width} = 5$ модель, обученная по методике Prod2Query, показывает максимальную точность 77.2%, и это значение находится на уровне SOTA-моделей. Динамика изменения $mAP@12$ для Prod2Query показывает, при $beam_{width} > 5$ точность снижается, и это соответствует предположению, что более длинные запросы не всегда приводят к ожидаемым результатам поиска.

IV. ЗАКЛЮЧЕНИЕ

Предложенная методика Prod2Query на основе генеративной модели машинного обучения показала себя эффективной для решения задачи холодного старта продаж на электронной торговой интернет-площадке. Модель, созданная в рамках Prod2Query, создает

синтетические данные о поведении пользователей относительно новых товаров. Как правило, вес поведенческого фактора в моделях машинного обучения для извлечения информации о товарах составляет 10-15%. Возможность управлять этим фактором через модель, а не дискретно на основе правил и шаблонов, позволит улучшить пользовательский опыт.

В промышленной эксплуатации модель, разработанная по методике Prod2Query, является моделью-учителем для моделей извлечения информации о товарах. Вклад от обученной с помощью Prod2Query модели извлечения товаров контролируем посредством инструментов платного продвижения товаров в выдаче, что является перспективой дальнейших исследований.

БИБЛИОГРАФИЯ

- [1] Bernardi L. et al. The continuous cold start problem in e-commerce recommender systems //arXiv preprint arXiv:1508.01177. – 2015.
- [2] Wang H. et al. A dnn-based cross-domain recommender system for alleviating cold-start problem in e-commerce //IEEE Open Journal of the Industrial Electronics Society. – 2020. – Т. 1. – С. 194-206.
- [3] Patro S. G. K. et al. Cold start aware hybrid recommender system approach for E-commerce users //Soft Computing. – 2023. – Т. 27. – №. 4. – С. 2071-2091.
- [4] Han C. et al. Addressing Cold Start in Product Search via Empirical Bayes //Proceedings of the 31st ACM International Conference on Information & Knowledge Management. – 2022. – С. 3141-3151.
- [5] Gupta P. et al. Treating cold start in product search by priors //Companion Proceedings of the Web Conference 2020. – 2020. – С. 77-78.
- [6] Gong Y. et al. An Unified Search and Recommendation Foundation Model for Cold-Start Scenario //Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. – 2023. – С. 4595-4601.
- [7] Missault P. et al. Addressing cold start with dataset transfer in e-commerce learning to rank. – 2021.
- [8] Li S. Embedding-based product retrieval in Taobao search // Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. — 2021. — С. 3181-3189.
- [9] Dai Z., Callan J. Deeper text understanding for IR with contextual neural language modeling //Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval. – 2019. – С. 985-988.
- [10] Nogueira R., Cho K. Passage Re-ranking with BERT //arXiv preprint arXiv:1901.04085. – 2019.
- [11] Dai Z., Callan J. Context-aware document term weighting for ad-hoc search //Proceedings of The Web Conference 2020. – 2020. – С. 1897-1907.
- [12] Nogueira R. et al. Document expansion by query prediction //arXiv preprint arXiv:1904.08375. – 2019.
- [13] Scells H., Zhuang S., Zuccon G. Reduce, reuse, recycle: Green information retrieval research //Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. – 2022. – С. 2825-2837.
- [14] Nogueira R., Lin J., Epistemic A. I. From doc2query to docTTTTTquery //Online preprint. – 2019. – Т. 6. – С. 2.
- [15] MacAvaney S. et al. Expansion via prediction of importance with contextualization //Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval. – 2020. – С. 1573-1576.
- [16] MacAvaney S., Tonello N., Macdonald C. Adaptive re-ranking with a corpus graph //Proceedings of the 31st ACM International Conference on Information & Knowledge Management. – 2022. – С. 1491-1500.
- [17] Mallia A. et al. Learning passage impacts for inverted indexes //Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. – 2021. – С. 1723-1727.
- [18] Zhuang S., Zuccon G. TILDE: Term independent likelihood moDEL for passage re-ranking //Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. – 2021. – С. 1483-1492.
- [19] Maynez J. et al. On faithfulness and factuality in abstractive summarization //arXiv preprint arXiv:2005.00661. – 2020.
- [20] Papineni K. et al. Bleu: a method for automatic evaluation of machine translation //Proceedings of the 40th annual meeting of the Association for Computational Linguistics. – 2002. – С. 311-318.
- [21] Doddington G. Automatic evaluation of machine translation quality using n-gram co-occurrence statistics //Proceedings of the second international conference on Human Language Technology Research. – 2002. – С. 138-145.
- [22] Lin C. Y. Rouge: A package for automatic evaluation of summaries //Text summarization branches out. – 2004. – С. 74-81.
- [23] Snover M. et al. A study of translation edit rate with targeted human annotation //Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers. – 2006. – С. 223-231.
- [24] Banerjee S., Lavie A. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments //Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization. – 2005. – С. 65-72.
- [25] Meister C., Vieira T., Cotterell R. If beam search is the answer, what was the question? //arXiv preprint arXiv:2010.02650. – 2020.
- [26] Zhao M., White M., Javed F. Query Rewrite for Low Performing Queries in E-commerce Based On Customer Behavior. – 2020.
- [27] Cho E. et al. Personalized search-based query rewrite system for conversational ai //Proceedings of the 3rd Workshop on Natural Language Processing for Conversational AI. – 2021. – С. 179-188.
- [28] Cui J. et al. Knowledge distillation across ensembles of multilingual models for low-resource languages //2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – IEEE, 2017. – С. 4825-4829.
- [29] Huang P. S. et al. Learning deep structured semantic models for web search using clickthrough data //Proceedings of the 22nd ACM international conference on Information & Knowledge Management. – 2013. – С. 2333-2338.
- [30] Nigam P. et al. Semantic product search //Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. – 2019. – С. 2876-2885.
- [31] Huang J. T. et al. Embedding-based retrieval in facebook search //Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. – 2020. – С. 2553-2561.
- [32] Klein G. et al. Opennmt: Open-source toolkit for neural machine translation //arXiv preprint arXiv:1701.02810. – 2017.
- [33] Post M. A call for clarity in reporting BLEU scores //arXiv preprint arXiv:1804.08771. – 2018.

Статья получена 22 апреля 2024. Ф.В.Краснов, Исследовательский центр ООО "ВБ СК" на базе Инновационного Центра Сколково. krasnov.fedor2@wb.ru, <http://orcid.org/0000-0002-9881-7371>.

Prod2Query: Solving the Problem of Cold Start for E-Commerce Using Generative Language Modeling

F.V. Krasnov

Abstract - Large online marketplaces introduce thousands of new items daily. In order to purchase a new item, users must locate it through a search function, and search engines now predominantly utilize behavioral indicators when indexing products, such as purchasing, clicking, and viewing items, which are unavailable for new products. This constitutes the issue of the "cold start" of sales. With the advent of generative language models, it is now feasible to train a language model on user behavior to generate search queries for novel products. Consequently, a collection of synthetic behavioral data for novel items is generated, which may be utilized to train a search engine. The primary aim of this study is to assess the extent to which the autonomous indicators of a search engine trained on such synthetic data for novel products improve. Prod2Query is founded on the architecture of an Encoder-Decoder model, based on BERT transformers. As a result of testing new products using the Prod2Query system, an indicator for the new product extraction model was obtained, with an mAP@12 score of 77.2%. This score is on par with state-of-the-art (SOTA) models, indicating that the "cold start" problem can be effectively addressed based on signals from sales representatives when introducing new products. By generating synthetic search queries and training search models using these signals, it is possible to achieve high levels of accuracy in product retrieval.

Keywords - LLM, cold start, product search, product retrieval systems, transformers, Information Retrieval.

REFERENCES

- [1] Bernardi L. et al. The continuous cold start problem in e-commerce recommender systems //arXiv preprint arXiv:1508.01177. – 2015.
- [2] Wang H. et al. A dnn-based cross-domain recommender system for alleviating cold-start problem in e-commerce //IEEE Open Journal of the Industrial Electronics Society. – 2020. – T. 1. – C. 194-206.
- [3] Patro S. G. K. et al. Cold start aware hybrid recommender system approach for E-commerce users //Soft Computing. – 2023. – T. 27. – №. 4. – C. 2071-2091.
- [4] Han C. et al. Addressing Cold Start in Product Search via Empirical Bayes //Proceedings of the 31st ACM International Conference on Information & Knowledge Management. – 2022. – C. 3141-3151.
- [5] Gupta P. et al. Treating cold start in product search by priors //Companion Proceedings of the Web Conference 2020. – 2020. – C. 77-78.
- [6] Gong Y. et al. An Unified Search and Recommendation Foundation Model for Cold-Start Scenario //Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. – 2023. – C. 4595-4601.
- [7] Missault P. et al. Addressing cold start with dataset transfer in e-commerce learning to rank. – 2021.
- [8] Li S. Embedding-based product retrieval in Taobao search // Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. — 2021. — C. 3181-3189.
- [9] Dai Z., Callan J. Deeper text understanding for IR with contextual neural language modeling //Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval. – 2019. – C. 985-988.
- [10] Nogueira R., Cho K. Passage Re-ranking with BERT //arXiv preprint arXiv:1901.04085. – 2019.
- [11] Dai Z., Callan J. Context-aware document term weighting for ad-hoc search //Proceedings of The Web Conference 2020. – 2020. – C. 1897-1907.
- [12] Nogueira R. et al. Document expansion by query prediction //arXiv preprint arXiv:1904.08375. – 2019.
- [13] Scells H., Zhuang S., Zuccon G. Reduce, reuse, recycle: Green information retrieval research //Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. – 2022. – C. 2825-2837.
- [14] Nogueira R., Lin J., Epistemic A. I. From doc2query to docTTTTTquery //Online preprint. – 2019. – T. 6. – C. 2.
- [15] MacAvaney S. et al. Expansion via prediction of importance with contextualization //Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval. – 2020. – C. 1573-1576.
- [16] MacAvaney S., Tonello N., Macdonald C. Adaptive re-ranking with a corpus graph //Proceedings of the 31st ACM International Conference on Information & Knowledge Management. – 2022. – C. 1491-1500.
- [17] Mallia A. et al. Learning passage impacts for inverted indexes //Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. – 2021. – C. 1723-1727.
- [18] Zhuang S., Zuccon G. TILDE: Term independent likelihood model for passage re-ranking //Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. – 2021. – C. 1483-1492.
- [19] Maynez J. et al. On faithfulness and factuality in abstractive summarization //arXiv preprint arXiv:2005.00661. – 2020.
- [20] Papineni K. et al. Bleu: a method for automatic evaluation of machine translation //Proceedings of the 40th annual meeting of the Association for Computational Linguistics. – 2002. – C. 311-318.
- [21] Doddington G. Automatic evaluation of machine translation quality using n-gram co-occurrence statistics //Proceedings of the second international conference on Human Language Technology Research. – 2002. – C. 138-145.
- [22] Lin C. Y. Rouge: A package for automatic evaluation of summaries //Text summarization branches out. – 2004. – C. 74-81.
- [23] Snover M. et al. A study of translation edit rate with targeted human annotation //Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers. – 2006. – C. 223-231.
- [24] Banerjee S., Lavie A. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments //Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization. – 2005. – C. 65-72.
- [25] Meister C., Vieira T., Cotterell R. If beam search is the answer, what was the question? //arXiv preprint arXiv:2010.02650. – 2020.
- [26] Zhao M., White M., Javed F. Query Rewrite for Low Performing Queries in E-commerce Based On Customer Behavior. – 2020.
- [27] Cho E. et al. Personalized search-based query rewrite system for conversational ai //Proceedings of the 3rd Workshop on Natural Language Processing for Conversational AI. – 2021. – C. 179-188.
- [28] Cui J. et al. Knowledge distillation across ensembles of multilingual models for low-resource languages //2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – IEEE, 2017. – C. 4825-4829.
- [29] Huang P. S. et al. Learning deep structured semantic models for web search using clickthrough data //Proceedings of the 22nd ACM international conference on Information & Knowledge Management. – 2013. – C. 2333-2338.
- [30] Nigam P. et al. Semantic product search //Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. – 2019. – C. 2876-2885.

[31] Huang J. T. et al. Embedding-based retrieval in facebook search //Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. – 2020. – C. 2553-2561.

[32] Klein G. et al. Opennmt: Open-source toolkit for neural machine translation //arXiv preprint arXiv:1701.02810. – 2017.

[33] Post M. A call for clarity in reporting BLEU scores //arXiv preprint arXiv:1804.08771. – 2018.