Aligning Ranking Objectives with E-commerce Search Intent

Dominic Seyler, Ehsan Ebrahimzadeh, Alex Cozzi and Abraham Bagherjeiran[1]

[1]*eBay Search Ranking and Monetization*

Abstract

Learning To Rank (LTR) from implicit user feedback is the predominant approach in large scale information retrieval systems for e-commerce. The standard LTR approaches that account for search intent in training ranking models focus primarily either on enriching the feature representation of the model by estimates of the search intent, or adopting an auxiliary intent prediction task in a multi-task learning setting, or debiasing the user feedback using intent-aware propensity models. In this work, we propose intent-aware LTR schemes by stratifying the training data with respect to intent groups identified by empirical labels that correspond to distinct item desirability distributions. Specifically, through importance sampling on training queries based on the richest engagement event attributed to the engaged item on the SRP, we train rankers that align the LTR objective with the corresponding item desirability distributions for the browse and purchase intents. In order to demonstrate the efficacy of the proposed training data stratification technique in generalizing across query segments with different underlying intents, we evaluate the rankers trained with different importance sampling weights on the traffic segment identified by the Browse search experience in a major e-commerce platform. In particular, we show that the ranker trained only on queries with post click conversion signals is significantly outperformed by a ranker that relies also on non-converting queries in training. We further demonstrate a higher variance in ranking efficiency in the traffic segment identified by the browse intent, due to the more severe distribution shift.

Keywords

query segmentation, learning to rank, e-commerce

1. Introduction

In e-commerce, a substantial portion of customer purchases are the result of searching and exploring the item inventory of the platform. This behavior is best described as exploratory search, characterized as a combination of exploratory browsing and focused searching [1, 2]. While focused search has been a primary research focus for the information retrieval community, user studies show that roughly 40% to 65% of users' goals are informational [3], implying that searchers are seeking novel items. It is therefore essential for industry scale information retrieval systems to adjust to the user need depending on the specific search scenario [4].

In this work, we study the effect of adjusting the label distribution in the Learning To Rank (LTR) training objective to the underlying intent in exploratory and focused search scenarios in an e-commerce setting. Specifically, we evaluate the generalization performance of the rankers

eCom'23: ACM SIGIR Workshop on eCommerce, July 27, 2023, Taipei, Taiwan

dseyler@ebay.com (D. Seyler); eebrahimzadeh@ebay.com (E. Ebrahimzadeh); acozzi@ebay.com (A. Cozzi); abagherjeiran@ebay.com (A. Bagherjeiran)

trained on varying label distributions across traffic segments based on the *search experience* in an e-commerce search engine. Previous work has accounted for capturing the user intent in LTR mainly by enriching the feature representations of the models. However, there are multiple advantages to dealing with user intent in a feature-agnostic setting, primarily due to the hardness of the intent prediction task in the absence of additional contextual information about the user's preferences, particularly in the early stages of the search journey.

We take an alternative intent-aware LTR approach, by stratifying the training data with respect to the label distribution that estimates the underlying item desirability distribution for different intent groups. Intent estimation policy can be a complex function of the user's query as well as the prior interactions in the search context, therefore, in our evaluations we identify the browse or focused search intent based on the type of search interaction experience. Specifically, we identify the browse intent based on the user traffic who explicitly adopt the *Browse experience* and explore the retrieved items in product category groups, in contrast to expressing a shopping intent by issuing a keyword query. We argue that due to the measurable differences in the post click conversion rates across the two search experiences, an LTR objective with a label distribution based on click engagements is more suitable for users in the exploration phase of their search journey; whereas an objective with an empirical label distribution based on purchases is more suitable for users with a focused search intent. We motivate the relationship between user engagement events (i.e., clicks and purchases) and underlying user intent (i.e., browse and focused search) by identifying distinct user behavior patterns in each segment, as surfaced in logged search data from a major e-commerce platform.

In our search traffic analysis, we observe that users who engage with items in Browse Result Pages (BRP) when browsing product categories are less likely to convert on the clicked items compared to the users who issue a keyword query and engage with the items on Search Result Pages (SRP). Therefore, we argue that browse queries based on a user-selected category drive more engagements for further exploration, and the ranking objective in this segment should be driven primarily by engagement efficiency. On the other hand, for focused queries with an underlying purchase intent, the ranking objective should be driven mostly by the conversion efficiency. Our study makes an initial step towards characterizing the fundamental sale/engagement efficiency trade-off in different traffic segments. This is achieved by training ranking policies on optimization objectives tailored to the corresponding intent class and serving requests from each traffic segment with the designated ranker for that class. Our work makes the following contributions:

- We propose training data stratification techniques for intent-aware LTR with intent strata identified by the empirical target labels that estimate distinct item desirability distributions in different search interaction scenarios.
- We identify distinct engagement patterns and measurable difference in post click conversion rates in e-commerce search interaction scenarios corresponding to browsing product category groups (i.e., exploration) versus expressing a purchase-driven intent (i.e., focused search).
- We evaluate the effectiveness of our intent-aware LTR methodology in an e-commerce ranking scenario by highlighting the variance in ranking efficiency in the intent segment identified by the browse experience, as the training label distribution of the ranking policy

changes.

2. Related Work

There is a remarkable body of work that focus on query intent prediction in the context of e-commerce [5, 6, 7, 8, 9, 10, 11]. While exploration has been identified as one of the major user intents in e-commerce settings [12, 13], incorporating exploratory search intent in the e-commerce ranking objective remains an open problem [14].

Once user's browsing/exploration intent is estimated, much less work has focused on how a search system can effectively support different user behavior. For instance, Rahman et al. [15] propose a conceptual recommendation framework to support exploratory search in e-commerce, by combining ranked results from a text based search engine with a recommender system based on user interactions. Ebrahimzadeh et al. [16] propose debiased estimators with intent-aware propensity estimates based on the number of clicks in a search context. Medlar et al. [17] classify users' search tasks in fine-grained categories, which fall within a behavioral spectrum between focused search and exploratory browsing. Rowley [18] focus on matching a browsing user to the correct product category, while the question of whether the ranking objective should vary depending on the user intent is not addressed. In our work, we proposes to train rankers for users with varying intent, where the preference for different user engagements is incorporated in the objective function.

Another line of work focuses on query understanding. For example, Wang et al. [19] parse a query into linguistic concepts that related to specific instances in a knowledge base. Dai and Callan [20] use contextual language models (such as Kenton and Toutanova [21]) in information retrieval to achieve a better understanding of the query context. Another related area of work focuses on a better semantic understanding of queries [22]. Applying these works to E-Commerce is not trivial and needs to address many challenges, such as vocabulary gap and data sparsity [14]. Furthermore, this particular research area does not clearly distinguish users by their intent and these methods are not applicable for category-based queries without any keywords. In contrast, our browse-specific rankers are optimized to support users with browse intent and can perform on keyword-less queries.

3. Methodology

3.1. Problem Setup

We consider a generic supervised LTR setting where we observe user feedback in the form of clicks on the SRP, as well as the subsequent post click transaction event attributed to the clicked items. A search $s \in \mathcal{S}$ is characterized by the query $q \in \mathcal{Q}$ made by some user $u \in \mathcal{U}$, the SRP that is an ordered collection of items (i.e., documents) $\mathcal{D}_s \in \mathcal{D}^N$ retrieved by the engine with respect to q as well as the user's click and post click purchase events, denoted by $c(\mathcal{D}_s) \in \{0, 1\}^N$ and $p(\mathcal{D}_s) \in \{0, 1\}^N$, respectively. It is standard to assume that ground truth relevance labels r_d are produced by some oracle according to an underlying item desirability distribution $\mathbb{P}_{\mathbf{r}_d|\mathbf{s}}(r_d|s)$ in the search context s. We learn a ranking policy by minimizing the statistical risk, corresponding to a search efficiency loss function, which is meant to approximate the expected number of user behavior events. Specifically, the statistical risk for a ranking function $f(d; s)$, which produces a score for each individual document $d \in \mathscr{D}_s$ given the search context s, is defined as

$$R(f) = \mathbb{E}[\mathscr{L}(f(\mathscr{D}_s), r(\mathscr{D}_s))], \tag{1}$$

where $\mathscr{L}(f(\mathscr{D}_s), r(\mathscr{D}_s)) = \ell(f(\mathscr{D}_s))^T r(\mathscr{D}_s)$ is the DCG-based ranking efficiency loss with respect to the ideal ranking defined with respect to $r(\mathscr{D}_s)$, suitably discounted by some function of the rank $\ell(f(\mathscr{D}_s)) = [-\lambda(rank_f(d))]_{d\in\mathscr{D}_s}$ attributed to the items via the scoring function f.

Since the underlying joint distribution of the search contexts and relevance labels is not known to the learner, the standard approach is to build an empirical risk estimate based on a sampled set $\mathcal{S}^\tau$ of search contexts with suitably defined empirical labels $\hat{r}(\mathscr{D}_s)$ to approximate the ground truth label distribution $\mathbb{P}_{\mathbf{r}_d|\mathbf{s}}(r_d|s)$ such that

$$\hat{R}(f) = \frac{1}{|\mathcal{S}^\tau|} \sum_{s\in\mathcal{S}^\tau} \mathscr{L}(f(\mathscr{D}_s), \hat{r}(\mathscr{D}_s)), \tag{2}$$

assuming all the training data from logged search contexts come from the same underlying distribution $\mathbb{P}_{\mathbf{s}}(s)\mathbb{P}_{\mathbf{r}_d|\mathbf{s}}(r_d|s)$. In this work, we are primarily interested in *estimation error* in our Empirical Risk Minimization (ERM) setup, oblivious to feature representation, hypothesis class, and training optimization scheme, which have to do with optimization and approximation error in the ERM setup.

3.2. Stratified Empirical Risk Minimization

We stratify the LTR training data based on intent groups $i \in \mathscr{I}$ identified by distinct item desirability distributions

$$\mathbb{P}_{\mathbf{r}_d|\mathbf{s},\mathbf{i}}(r_d|s, i) \tag{3}$$

in different search interaction scenarios. We can therefore build stratified empirical risk estimates by adjusting the empirical labels $\hat{r}_i(\mathscr{D}_s)$ so that they approximate the underlying desirability distribution in the corresponding intent segment; that is

$$\hat{R}(f) = \sum_{i\in\mathscr{I}} \frac{\hat{\mathbb{P}}_i}{|\mathcal{S}_i^\tau|} \sum_{s\in\mathcal{S}_i^\tau} \mathscr{L}(f(\mathscr{D}_s), \hat{r}_i(\mathscr{D}_s)), \tag{4}$$

where $\mathcal{S}_i^\tau$ is the logged training data corresponding to the intent segment i and $\hat{\mathbb{P}}_i$ is some stratification distribution across the intent groups. We focus on a simple binary intent segmentation based on the underlying browse or purchase behavior of the user. In this case, our estimate $\hat{r}_i(\mathscr{D}_s)$ of the underlying intent-based item desirability distribution $\mathbb{P}_{\mathbf{r}_d|\mathbf{s},\mathbf{i}}(r_d|s, i)$ is two-fold: for users with the browse intent it is the empirical click engagement distributions $c(\mathscr{D}_s)$, whereas for users with focused search intent it is estimated the empirical distribution based on purchase signals $p(\mathscr{D}_s)$. In fact, to avoid the complexity in building our empirical estimate, let us assume that we only have a single clicked item within all the training search contexts. Thus, since we observe post-click purchases only for clicked items in converting queries, our stratified estimate can be written as

$$\hat{R}(f) = \frac{1}{|\mathcal{S}^{\tau}_{\alpha}|} \sum_{s \in \mathcal{S}^{\tau}_{\alpha}} \mathcal{L}(f(\mathcal{D}_s), \hat{c}(\mathcal{D}_s)), \tag{5}$$

where $\hat{c}(\mathcal{D}_s))$ is the suitably debiased click label(e.g. based on inverse propensity weighting) and the training search contexts $\mathcal{S}^{\tau}_{\alpha}$ are sampled from the search context with post click purchases $\mathcal{P}$with an importance sampling parameter α and from the search context with no post-click purchases $\mathcal{C}$ with importance sampling parameter $1-\alpha$; that is $\mathcal{S}^{\tau}_{\alpha} = \{\mathcal{P}\}_{\alpha} \cup \{\mathcal{C}\}_{(1-\alpha)}$.

Using standard arguments, we can characterize the bias and variance of the proposed stratified empirical risk estimator with respect to the Bayes statistical risk corresponding to the underlying ground truth item desirability distribution in the segment. In fact, the generalization power of the proposed stratified empirical risk estimate can be adjusted to the corresponding query segment by an appropriate choice of the stratification parameter α.

4. Experimental Evaluation

4.1. Evaluation Segments

While we can clearly highlight the generalization power of the proposed stratified risk estimation technique by focusing on the same strata on the evaluation dataset, we instead rely on evaluation segments that can explicitly be identified at the online serving time. Specifically, we identify the user's underlying search intent based on the the explicit search engine interaction experience that the user opts into: (A) the user expresses a browse intent by clicking on a category on the homepage, e.g. "Automotive", and explores the results on a Browse Result Page (BRP); (B) the user expresses a focused search intent by issuing a query by entering keywords in the search bar, e.g. "Audi R8 headlights", and explores the results on a Search Result Page (SRP).

To motivate the relationship between user intent, as characterized by BRPs and SRPs, with the corresponding intent strata at the training time, we perform an analysis of post click conversion rates (CVR) on user search logs. In fact, we report a measurably lower CVR in BRPs compared to SRPs, with a difference $\Delta(CVR_{SRP}, CVR_{BRP}) > 50\%$. This is a strong indicator that BRP engagements are more likely to be followed by subsequent clicks via engaging in other item recommendation modules or issuing a keyword search query rather than committing to a purchase. This is a key observation as it motivates the connections between the intent classes and the corresponding engagement/conversion driven objectives. To further understand the interaction patterns across query intent segments, we perform a second study to compare click propensities at different ranks across browse and focused search query segments (Figure 1). We observe a smoother rank-decay in click propensities in browse queries and more engagement likelihood in lower ranks. This observation confirms the hypothesis that user behavior patterns vary by intent, as users with a browse intent are more likely to explore lower ranks, while users with a purchase intent are more likely to engage with top ranks.

The inherent differences in the engagement patterns between browse and focused search queries indicate that: (1) Browse queries drive engagements for further exploration, thus the ranking objective in this segment should be driven by the engagement efficiency; and (2) In

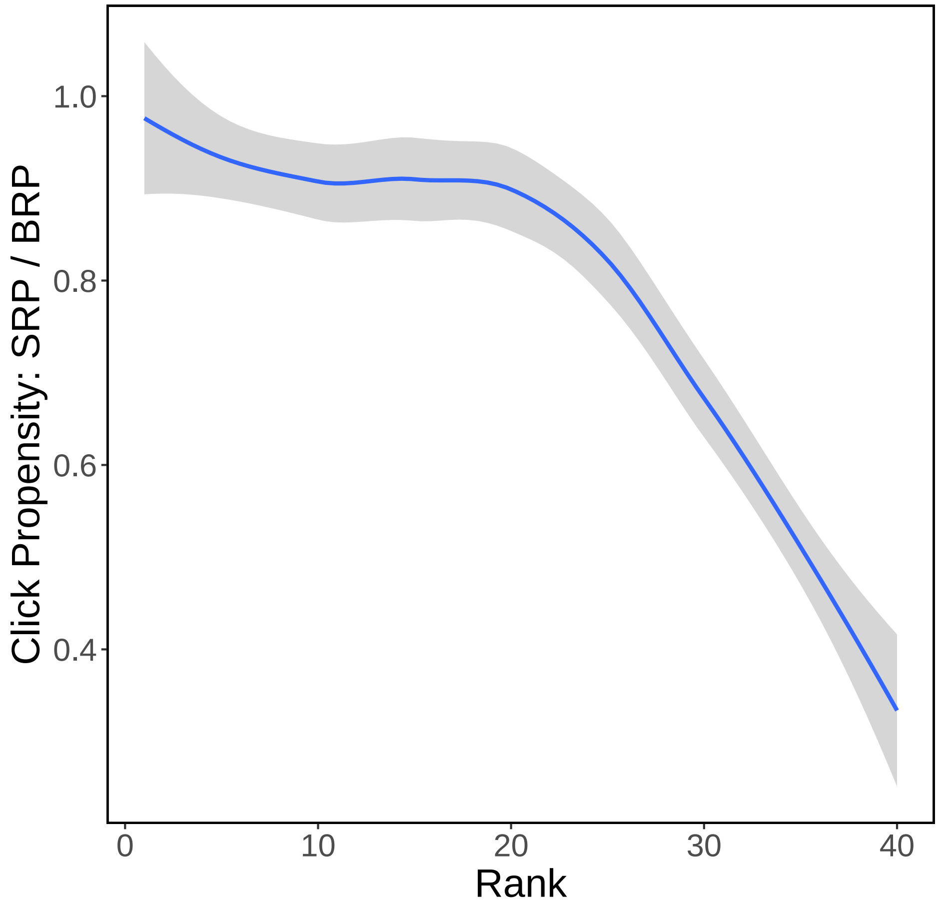

Figure 1: Click Propensity Ratios of Search Result Page (SRP) and Browse Result Page (BRP) across Different Ranks.

contrast, focused search queries are more likely to be specific and purchase driven, thus the ranking objective should primarily be driven by efficiency of conversion events.

4.2. Experiment Setup

In this section we design our experiments to measure the empirical performance of the rankers trained on segment-specific training objectives evaluated on query segments corresponding to BRP and SRP search experiences. Given an evaluation segment, we are specifically interested in the *relative performance* of a ranker that is trained only on queries with post click purchases attributed to the engagement events versus one that can rely also on the non-converting search contexts. We train ranking models f_α, parameterized by the importance sampling parameter α, on stratified training datasets $\mathcal{S}_\alpha^\tau$, where α proportion of the training queries are sampled from queries with a post click purchase event and the remaining $1 - \alpha$ proportion are sampled from the non-converting query set. By training different rankers f_α through varying the stratification parameter α, and evaluating them over different user intent segments, we establish a conversion/engagement trade-off as it relates to the underlying item desirability label distribution of the ranker.

For a logged search context s in the evaluation dataset $\mathcal{S}^{\epsilon}$, suppose that the richest engagement event attributed to the engaged item on the SRP is E; which is essentially a purchase event P if there is a post-click conversion attributed to that engaged item, otherwise it is just the click event C on the SRP. The ranking efficiency of the ranker f is measured as the average Mean Reciprocal Rank (MRR) with respect to $rank_f(E)$, the rank attributed to the engaged item by f in search contexts $s \in \mathcal{S}^{\epsilon}$; that is

$$MRR_E(f, \mathcal{S}^{\epsilon}) = \frac{1}{|\mathcal{S}^{\epsilon}|} \sum_{s \in \mathcal{S}^{\epsilon}} \frac{1}{rank_f(E)}. \tag{6}$$

4.3. Datasets

Training Data. We build distinct datasets by stratifying the queries based on the richest engagement event observed in the search context $c(\mathcal{D}_s)$ and $p(\mathcal{D}_s)$; Specifically, the proportion of queries with a purchase event in the search context is controlled by importance sampling parameter α. Note that during training, models are exposed to queries from all intent segments (both in terms of the training segments based on the underlying item desirability distribution and the evaluation segments based on the BRP/SRP search experience) to take advantage of synergies among all user experiences. To control for the size of the training dataset, for each query we sample three unengaged items from the candidate set $\mathcal{D}_s$ at random, as negative examples. For training targets $\hat{c}(\mathcal{D}_s)$, we use debiased labels using the propensities characterized in the previous section. All training datasets contain roughly 1M queries.

Test Data. The test datasets are based on a random sample of queries for which there is at least one click event observed in the search context, suitably separated from training sets to avoid data leakage. The two test datasets are built via stratification based on BRP and SRP intent segments and each contain around 10K to guarantee meaningfully narrow confidence intervals. For the test data, we keep all the candidate items to be re-ranked by the candidate ranker.

Note that datasets are a random sample of e-commerce search traffic in the months of April to June 2022.

4.4. Empirical Results

Table 1 offers a fundamental observation on the generalization power of the proposed stratified empirical risk estimate. Essentially, we can achieve a bias/variance trade-off by an appropriate choice of the stratification parameter α, which can in fact be adjusted based on the evaluation segment that ranker is served on. As such, the unbiased estimator of the purchase desirability intent $f_{\alpha=1}$, fails to generalize even on the traffic population with purchase intent, primarily due to the selection bias in training search context qualification. Essentially, $f_{\alpha=1}$ relies only on converting queries to estimate the more diverse traffic distribution. This is particularly more pronounced in the BRP segment, because as pointed out in the search traffic analysis, most users do not commit to a purchase upon engaging with an item in a BRP. In fact, by relying also on non-converting queries in training the rankers via stratified empirical risk estimates, we can improve the estimation accuracy of the target distribution for engagement efficiency.

Table 1
Performance difference between purchase-driven model $f_{\alpha=1}$ and the baseline model with optimal $f_{\alpha*}$. All reported lifts are statistically significant and exceed the threshold for a meaningful effect size

Lift w.r.t Evaluation Data	$\mathcal{S}^{\epsilon}_{BRP}$	$\mathcal{S}^{\epsilon}_{SRP}$
$\Delta\,(MRR_C(f_{\alpha=1},\cdot), MRR_C(f_{\alpha*},\cdot))$	-8.3%	-5.7%
$\Delta\,(MRR_P(f_{\alpha=1},\cdot), MRR_P(f_{\alpha*},\cdot))$	-2.1%	-0.7%

On the other hand, if we do not over-sample converting queries and only rely on a randomized sample from the traffic population $\mathcal{S}^C$, both the conversion and engagement rates will be meaningfully impacted, even in population segments with primarily a browse intent, because search contexts with post click purchase $\mathcal{S}^P$ usually offer richer information for training.

Next, by comparing the effect size of the relative performance of the rankers across query segments, we highlight a meaningfully *higher variability in engagement efficiency* in the browse segment compared to the focused query segment, highlighting strong evidence towards the choice of a click engagement driven objective for the browse segment. In fact, we report a significantly higher variance in the search efficiency metric MRR_C in the BRP segment compared to the SRP segments as the ranking model f_α varies; that is

$$\Delta\left(\underset{\alpha}{\mathrm{Var}}[MRR_C(f_\alpha, \mathcal{S}^{\epsilon}_{BRP})], \underset{\alpha}{\mathrm{Var}}[MRR_C(f_\alpha, \mathcal{S}^{\epsilon}_{SRP})]\right) > 10\% \tag{7}$$

The large variance difference is due to a significant performance drop for model f_α when α is close to 1, as they fail to generalize to the population distribution with an underlying browse intent. In fact, the best performing setting for α can be tuned optimally for the specific segments that ranking policy serves. We leave the exploration of other query segmentation strategies, beyond BRP and SRPs, as well as the best practice for fine tuning the stratification parameter to future work.

5. Conclusion

We presented a training data stratification technique based on intent groups characterized by distinct underlying item desirability distributions. To verify the generalization performance of the proposed scheme, we performed empirical evaluations on search traffic from a major e-commerce platform. We showed the poor generalization of the ranker trained only on queries with rich post click engagement events, and demonstrated that as we change the training target distribution of the ranker, the variance in search efficiency is measurably higher in the browse segment compared to the traffic segment identified by keyword queries. These observations confirm the hypothesis on the fundamental benefits of aligning the ranking objective with the user's search intent.

References

[1] M. J. Bates, What is browsing-really? a model drawing from behavioural science research, 2007.

[2] G. Marchionini, Exploratory search: from finding to understanding, Communications of the ACM 49 (2006) 41–46.
[3] J. Teevan, C. Alvarado, M. S. Ackerman, D. R. Karger, The perfect search engine is not enough: a study of orienteering behavior in directed search, in: Proceedings of the SIGCHI conference on Human factors in computing systems, 2004, pp. 415–422.
[4] T. Ruotsalo, J. Peltonen, M. J. Eugster, D. Głowacka, P. Floréen, P. Myllymäki, G. Jacucci, S. Kaski, Interactive intent modeling for exploratory search, ACM Transactions on Information Systems (TOIS) 36 (2018) 1–46.
[5] L. Guo, L. Hua, R. Jia, B. Zhao, X. Wang, B. Cui, Buying or browsing?: Predicting real-time purchasing intent using attention-based deep network with multiple behavior, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, pp. 1984–1992.
[6] D. Cirqueira, M. Hofer, D. Nedbal, M. Helfert, M. Bezbradica, Customer purchase behavior prediction in e-commerce: A conceptual framework and research agenda, in: New Frontiers in Mining Complex Patterns: 8th International Workshop, NFMCP 2019, Held in Conjunction with ECML-PKDD 2019, Würzburg, Germany, September 16, 2019, Revised Selected Papers, Springer, 2020, pp. 119–136.
[7] B. Zheng, B. Liu, A scalable purchase intention prediction system using extreme gradient boosting machines with browsing content entropy, in: 2018 IEEE International Conference on Consumer Electronics (ICCE), IEEE, 2018, pp. 1–4.
[8] P. Minjing, L. Xinglin, L. Ximing, Z. Mingliang, Z. Xianyong, D. Xiangming, W. Mingfen, Recognizing intentions of e-commerce consumers based on ant colony optimization simulation, Journal of Intelligent & Fuzzy Systems 33 (2017) 2687–2697.
[9] D. Schellong, J. Kemper, M. Brettel, Generating consumer insights from big data clickstream information and the link with transaction-related shopping behavior (2017) 24. URL: http://aisel.aisnet.org/ecis2017_rp/24.
[10] D. Schellong, J. Kemper, M. Brettel, Clickstream data as a source to uncover con-sumer shopping types in a large-scale online setting, in: 24th European Conference on Information Systems, ECIS 2016, Istanbul, Turkey, June 12-15, 2016, 2016, p. Research Paper 1. URL: http://aisel.aisnet.org/ecis2016_rp/1.
[11] M. Zavali, E. Lacka, J. De Smedt, Shopping hard or hardly shopping: Revealing consumer segments using clickstream data, IEEE Transactions on Engineering Management (2021).
[12] N. Su, J. He, Y. Liu, M. Zhang, S. Ma, User intent, behaviour, and perceived satisfaction in product search, in: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, 2018, pp. 547–555.
[13] P. Sondhi, M. Sharma, P. Kolari, C. Zhai, A taxonomy of queries for e-commerce search, in: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, 2018, pp. 1245–1248.
[14] M. Tsagkias, T. H. King, S. Kallumadi, V. Murdock, M. de Rijke, Challenges and research opportunities in ecommerce search and recommendations, in: ACM SIGIR Forum, volume 54, ACM New York, NY, USA, 2021, pp. 1–23.
[15] M. M. Rahman, Z. M. Zaki, N. H. B. M. Alwi, M. M. Islam, A conceptual model for the e-commerce application recommendation framework using exploratory search, Solid State Phenomena (2020).

[16] E. Ebrahimzadeh, A. Cozzi, A. Bagherjeiran, Intent-aware propensity estimation via click pattern stratification, in: Companion Proceedings of the ACM Web Conference 2023, 2023, pp. 751–755.
[17] A. Medlar, J. Pyykkö, D. Glowacka, Towards fine-grained adaptation of exploration/exploitation in information retrieval, in: Proceedings of the 22nd International Conference on Intelligent User Interfaces, 2017, pp. 623–627.
[18] J. Rowley, Product search in e-shopping: a review and research propositions, Journal of consumer marketing (2000).
[19] Z. Wang, K. Zhao, H. Wang, X. Meng, J.-R. Wen, Query understanding through knowledge-based conceptualization, in: Twenty-Fourth International Joint Conference on Artificial Intelligence, 2015.
[20] Z. Dai, J. Callan, Deeper text understanding for ir with contextual neural language modeling, in: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2019, pp. 985–988.
[21] J. D. M.-W. C. Kenton, L. K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of NAACL-HLT, 2019, pp. 4171–4186.
[22] R. Baeza-Yates, Semantic query understanding, in: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2017, pp. 1357–1357.