Paper Review: Mitigating Position Bias in Hotels Recommender Systems

BY: Yinxiao Li

2.1)

Context

Recommendation Systems and ranking algorithms in search are cornerstones in today's digital environment for users to reach their targeted results on numerous platforms. Most of these recommendation systems and ranking algorithms depend upon large volumes of implicit feedback such as clicks, conversions, and dwell times from the users themselves to finetune rankings for optimization. These data can be biased; one such bias that is normally prevalent in these data is position bias. This bias arises because users tend to click on items that come top in the list, either by convenience or by trusting the ranking of the platform. The present paper explores this issue in the context of TripAdvisor Hotel recommendations. The writer seeks to appeal to the need for ranking models that are true user interest and unbiased, rather than just positional placement, which becomes highly crucial when there is a long list of items in the system, as would occur in hotel booking systems where users might never examine every option.

Problems

These issues include position bias: users tend to click on high-ranked items more frequently. The data, hence, is biased and representative not of the actual preference of the users. How these biases may be dealt with in various ways, by discarding all data beyond the last clicked position or by taking into consideration all the results with their weights-is limited in such a way that one method loses valuable feedback, while another one unnecessarily overcomplicates the ranking models. Most of the click propensity estimation methods likelihood based on items' position rely on randomized interventions in user interfaces, which hurts the user experience. This problem is even harder in dynamic settings like hotel recommendations external factors like hotel pricing further complicate getting a reliable and unbiased ranking model.

Contributions

The paper develops a novel method that incorporates the best of previous models for reducing position bias without interface randomization. The writers develop a propensity sampling method that estimates click likelihood given user click data in this paper. The work involves the creation of a dataset that retains all entries up to the last clicked position while selectively including lower-ranked items based on estimated observation likelihoods. It follows a pairwise ranking model that is trained on pairs of data for efficient relative relevance. Experimentation through A/B testing using live data on TripAdvisor shows a propensity-sampled model results in a 1.5% increase in clicks, indicative of improved engagement and effective mitigation of position bias. That being said, this makes the approach more valuable for the research community because it offers a scalable, user-centric solution that can better recommendations without extensive reconfiguration on increased model complexity.

2.2) Critically evaluate

What is the type of paper?

This paper focuses on position bias in hotel recommendation systems and is presented by Yinxiao Li. It can be termed as an applied research paper since it tries to solve a very real problem in the recommendation system: the ranking of hotel search results based on user behavior biases. While the paper goes in-depth into a theoretical investigation of position bias and its related biases, it also applies practical methods such as propensity sampling and unbiased Learning-to-Rank frameworks to develop and verify a ranking model using A/B testing for its practical relevance and impact.

How well are the research questions addressed?

The paper addresses the critical research question of position bias mitigation in hotel recommendation systems-one of the key challenges in Learning-to-Rank models in which users are likely to click higher-ranked results without consideration for relevance and, correspondingly, generate biased data that inappropriately lead to suboptimal rankings.

It evaluates three existing methodologies: (1) Naïve Inclusion of All Data, which assumes users evaluate all results equally but fails to correct biases; (2) Truncation at the Last Click, which retains only results up to the last clicked item but introduces systematic biases, such as prioritizing rank reversal; and (3) Unbiased Loss Function Using Propensity Weights, which uses weighted loss functions for debiasing but requires accurate propensity estimation and overlooks user certainty for higher-ranked results. It further proposes a hybrid methodology, combining strengths from the second and third approaches. This hybrid approach uses user action data for propensity sampling but discards irrelevant data.

This model excludes search logs without clicks or bookings; results above or equal to the lowest clicked position are retained unsampled, while those lower than that position are sampled based on estimated propensities. First, propensities are calculated based on historical booking data in a very robust and user-friendly manner, avoiding randomization experiments that could be intrusive. The hybrids' performances are validated via A/B testing in the live system of TripAdvisor, yielding a relative improvement in CTR of 1.5% compared to the baseline using 100% sampling and of 1.7% compared to a model using 80% sampling. This hybrid model effectively avoids position bias, enhances the user's interaction metrics, and ensures scalability and feasibility for real-world deployment by integrating theoretical foundations and practical applications. The paper perfectly bridges the gap between theoretical understanding and practical solutions to come up with a robust, efficient, and user-centric approach to solving position bias in recommendation systems.

Is the approach realistic?

It has been quite a realistic and pragmatic approach. The paper has evinced an awareness of what the traditional methods lacked, such as the degradation of user experience owing to randomization experiments, and steered clear of these pitfalls by using propensity-based adjustments derived from historical data. Lightweight modifications to the existing infrastructure made the proposed solution feasible for widescale deployment in a commercial setting like TripAdvisor. Additional practical constraints involve dynamic hotel prices and how they might influence user behavior to make the model as realistic as possible. Besides this, the scalability and generalizability of the method are supported by showing its application in an online A/B testing setup that meaningful performance improvements without complicating the system.