Trigger Relevancy and Diversity Inefficiency with Dual-Phase Synergistic Attention in Shopee Recommendation Ads System

Sheng Li*
sheng.li@shopee.com
Shopee Pte. Ltd.
Singapore

Zhiwei Chen*
zhiwei.chen@shopee.com
Shopee Pte. Ltd.
Singapore

Abstract

Shopee stands as the foremost e-commerce platform across Southeast Asia and Latin America, where like its counterparts, it utilizes deep learning techniques to enhance its ads recommendation systems. Different types of ads recommendation scenarios emerged to satisfy the needs of different users. Among these, Trigger-Induced Recommendation (TIR) is a relatively new field that has recently gained attention for targeted optimizations from both the industry and the research field. In this work, we pinpointed a problem specific to TIR scenarios in the industry, named Trigger Relevancy and Diversity Inefficiency (TRDI). Illustrating with Shopee Product Detail Page You May Also Like (PDP YMAL) recommendation section, where billions of item impressions occur daily, we introduce a systematic approach for the industry community to examine the extent of TRDI in their TIR scenarios. Then we propose a novel approach: Dual-Phase Synergistic Attention (DPSA) method, to tackle the TRDI problem. We proved the effectiveness of DPSA by integrating it with Shopee PDP YMAL recommendation section's Ads CTR ranking model. After thorough online A/B testing and holdout testing, DPSA has become the new baseline at Shopee PDP YMAL PaidAds CTR ranking across all regions of Shopee markets, significantly enhancing key metrics for user experience, seller profits, and platform gains.

CCS Concepts

• **Information systems** → **Online advertising**; **Personalization**; **Content ranking**; **Retrieval models and ranking**.

Keywords

online advertising, recommendation, trigger-induced recommendation, user-behavior modeling

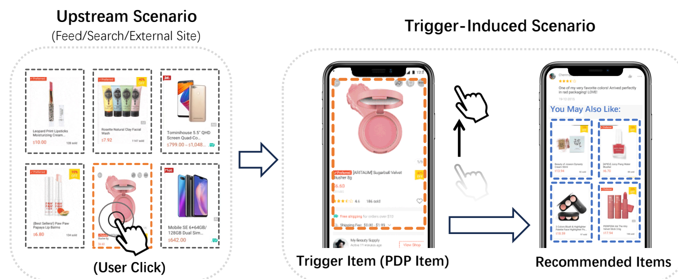

Figure 1: Entering the Product Detail Page (PDP) from any scenario and scrolling down leads into the YMAL (You May Also Like) scenario, a typical Trigger-Induced Scenario.

1 Introduction

In Shopee, online shopping recommendation scenarios could be broadly categorized into two types: 1) Exploratory Recommendation, such as Home Page Daily Discover, where recommended items are primarily attributed to user interests inferred from historical interactions and item's platform popularity; 2) Trigger-Induced Recommendation (TIR) [1–3], where a trigger item activates the follow-up recommendations. TIR is a specialized recommendation scenario that recently gained attention from both the industry and research field, calling for more targeted optimization efforts. Compared to exploratory recommendations, TIR has an additional purpose to fulfill by product design, which is to satisfy users' need to conveniently view similar or complementary (trigger-relevant) to the trigger item. Taking Shopee PDP YMAL in Figure 1 for instance, when clicking on any item card, a user will enter the Product Detail Page (PDP), where the detailed information of the clicked item will be displayed, and the clicked item is then regarded as the PDP item. When scrolling down the PDP, there is an ads recommendation section named You May Also Like (YMAL), where the PDP item acts as the trigger item to induce recommendations. As Shopee PDP YMAL is the largest ads recommendation entrance in Shopee, it contributes the highest amount of impressions, conversions, and revenue among all recommendation entrances, thereby calling for investments in targeted optimization efforts.

While the initial product motivation of setting up TIR scenario in the industry often begins with offering an entrance for users to discover similar or complement items (trigger-relevant) with relative ease, we have observed in Shopee PDP YMAL that recommendations not relevant to the trigger item (trigger-diverse) but potentially interesting are also receiving a fair share attention from our users. Such observation induced us to investigate if there exists inefficiency in the allocation of trigger-relevant and trigger-diverse recommendations by our recommendation system. With careful analyses from different perspectives in Shopee PDP YMAL, we identified **Trigger Relevancy and Diversity Inefficiency (TRDI)**. Take an extreme illustration of such inefficiency: a user only clicks

*Both authors contributed equally to this research.

AdKDD'24, August 26, 2024, Barcelona, Spain

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

on trigger-diverse recommendations when browsing Shopee PDP YMAL, but our recommendation system fails to recognize such user preference and still displays many trigger-relevant recommendations to this user. Conversely, the same is true. This results in low traffic efficiency in the recommendation pipeline. We believe such inefficiencies may also prevail in the TIR scenarios of other industry players, such as YouTube's next video recommendations and Amazon's You may also like recommendations. In this paper, we share our systematic approach to assess the extent of TRDI in Shopee PDP YMAL from different perspectives namely *individual-level* and *trigger-item-level*, which could be beneficial for other industry players in assessing their TIR scenarios. To the best of our knowledge, there is no prior work focusing on TRDI.

In this work, we introduce two new terms: trigger-relevancy and trigger diversity. The understandings of relevancy and diversity in the general recommendation are to be distinguished from these two terms. Recommendation relevancy is often referred to as the accuracy of recommendations, which is performance-related. Several classic information retrieval metrics are often used to gauge the relevancy performance of recommendation algorithms. For instance, the recommendation ranking phase often uses metrics like NDCG, AUC, or GAUC to measure the discriminative power of ranking algorithms by ranking more relevant items to the front of a sorted list of recommendation items. Trigger-relevancy, on the other hand, is not performance-oriented, it simply refers to the relatedness to the trigger item, which can be preliminarily characterized by simple rules such as whether the recommended item has the same category as the trigger item. In other words, trigger-relevant recommendations are similar to complementary to the trigger item. On the other hand, diversity in recommendation generally focuses on the problem of users suffering *redundancy issue* of viewing redundant and similar items repeatedly within a time window [4–6]. Diversity recommendation is therefore a complicated list-wise optimization problem that handles the interference among different displayed recommendations within the whole system, or within a specific period for a user. Trigger-diversity, however, simply refers to recommended items that are not related to the trigger item. These items are often recommended because of their platform popularity (trending) or because the user has shown interest in similar items previously.

In recognition of TRDI, we propose Dual-Phase Synergistic Attention (DPSA) to mitigate TRDI. DPSA, as a standalone module, can be easily fused with any deep learning model used in typical recommendation phases including recall, prerank, and ranking. Compared to the conventional Clicked Item Sequence $Seq^{(c)}$ used in many classic recommendation frameworks like [7–10], DPSA introduces a one-to-one corresponding relationship between $Seq^{(c)}$ and Trigger Item Sequence $Seq^{(t)}$. In the first phase of DPSA, we borrowed the idea from the additive attention mechanism used by [7, 11] to calculate a relevancy preference score for each user, in hope to mitigate the individual-level issue of TRDI. The second phase of DPSA extensively explores the relationship between $Seq^{(c)}$ and $Seq^{(t)}$, capturing users' behavioral patterns under different trigger items, thereby tackling the trigger-item-level issue of TRDI. We conducted extensive offline and online experiments with Shopee PDP YMAL PaidAds CTR ranking model to verify the effectiveness of DPSA. The main contributions of this paper are summarized as follows:

- Begin with Shopee PDP YMAL, we have, for the first time in the industry, identified the TRDI issue in TIR from individual-level and trigger-item-level. We believe such a pattern is not unique to Shopee PDP YMAL; its presence in other TIR scenarios across the industry should be anticipated. The extent of TRDI could be measured with the approach we described in this paper.
- We proposed DPSA, which can be easily integrated into existing recommendation frameworks in many stages of the system (recall, pre-rank, rank, etc.) to improve the personalizing ability in TIR scenarios.
- We have proven the effectiveness of the DPSA in the Shopee PDP YMAL advertising scenario through both offline and online experiments focusing on CTR prediction in the ranking phase. The implementation was subsequently released as the production model for the Shopee PDP YMAL PaidAds CTR ranking model in all regions, achieving +1.12% CTR, +2.10% CTCVR, +1.78% Revenue, and +3.32% GMV.

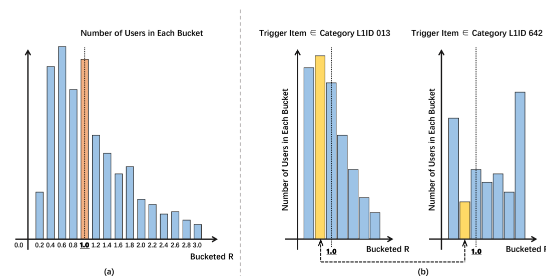

Figure 2: The distribution of R scores across users, where the horizontal axis represents the bucketed R scores, and the vertical axis indicates the number of users in each bucket.

2 Related Works

Trigger-Induced Recommendation (TIR) is a relatively new research area that has been gaining attention over the past few years. Besides, user behavior modeling is also a key component in this work. In this section, we review related works in these two areas.

Trigger-Induced Recommendation (TIR). [3] first proposed a novel framework named R3S to solve a recommendation suggestion task for extended reading in recommendation. R3S decides whether to display relevant items for extended reading, and provides relevant items if the decision to display is confirmed. Follow-up works by [1, 2] attempted to formalize such recommendation task as a special kind of recommendation. [1] named such a recommendation scenario as *Relevant Recommendation*, where relevant items are recommended when users express interest in one target item. On the other hand, [2] named such recommendation scenario as *Trigger-Induced Recommendation (TIR)*, where users' instant interest could be explicitly represented by a trigger item, such that the follow up recommendations are induced by a trigger item. In this paper, we follow the terminology of TIR as proposed by [2]. These prior works formalized TIR as a special kind of recommendation scenario in the industry. In this work, we identify the Trigger Relevancy and Diversity Inefficiency (TRDI) issue specifically for TIR scenarios and present its significance in Shopee.

User behavior Modeling User behavior modeling studies users' past behaviors to extract useful information or characteristics to enhance the recommendation efficiency. As Attention is All you Need was published [12] and started the revolution in the NLP realm to understand the natural text as a sequence, the recommendation realm has also started studying the application of attention variations on user behavior sequence modeling. For example, the classic DIN [7] presented a preliminary usage of additive attention mechanism [11] to capture user's diverse interests from rich historical behaviors; DMT [8] argued that different types of user behavior sequence (clicks, add-to-cart, order, etc) should have different levels of influence in the final ranking predictions, and the extent of weights could be learned and personalized for each user; SIM [10] proposed a search-based interest modeling technique that enables modeling of much longer sequence of user behaviors. In this work, specific to the unique TRDI issue identified for TIR scenario at Shopee PDP YMAL, we present a two-phase attention module named Dual-Phase Synergistic Attention (DPSA) to better model user behavior sequence, thereby enhancing Shopee's recommendation efficiencies.

3 Methodology

3.1 Examine TRDI

In our approach, trigger-relevant is preliminarily characterized by having the same coarse category as the trigger item, and vice versa for trigger-diverse. To ensure confidence in the analysis, only users with significant exposure (at least 100 impressions) to both trigger-relevant and trigger-diverse items in Shopee PDP YMAL in 14 days are considered. For each user, a score R is calculated, representing the ratio of the CTR of impressed trigger-relevant items to that of impressed trigger-diverse items. When R is greater than 1.0, it is reasonable to infer that the user prefers trigger-relevancy, and trigger-diverse items may have been recommended more than necessary, which implies inefficiency in TIR scenarios like Shopee PDP YMAL. Vice versa when R is lower than 1.0, implying a user's propensity towards trigger-diverse recommendations is not efficiently satisfied when viewing recommended items.

After computing R scores for all users with sufficient exposure to both trigger-relevant and trigger-diverse recommendations at Shopee PDP YMAL, the R scores distribution of these users is further examined from two levels: **1) Individual-level:** Figure 2 (a) shows that the majority of users have their R scores deviated from 1.0, which not only indicates the propensity to trigger-relevancy and trigger-diversity indeed differs across individual users but also implies that many users' preference for trigger-relevant or trigger-diverse recommendations are not efficiently satisfied. **2) Trigger-item-level:** Zoom into specific coarse categories of trigger item in Figure 2 (b), two observations are made. Firstly, the distributions of R scores differ across different categories of trigger items, implying different categories of items have different general likelihoods of receiving positive feedback from the users when recommended as trigger-relevant recommendations or trigger-diverse recommendations. Secondly, even within the same category, different users' propensity for trigger-relevancy or trigger-diversity could vary.

The discovery of differing R score distributions from individual-level and trigger-item-level suggests inefficiency in our recommendation system at Shopee PDP YMAL. We name such inefficiency Trigger Relevancy and Diversity Inefficiency (TRDI). We recommend industry players perform similar analyses to assess the extent of TRDI in their TIR scenarios.

3.2 Dual-Phase Synergistic Attention (DPSA)

Dual-Phase Synergistic Attention (DPSA) illustrated in Figure 3 consists of two phases of attention, each designed to tackle the individual-level and trigger-item-level TRDI problem identified in Section 3.1.

3.2.1 *First Phase of DPSA*

In light of the *individual-level TRDI*, where it reveals that propensity to trigger-diversity differs across individual users, the first phase of DPSA aims to personalize trigger-relevancy and trigger-diversity.

We borrowed the idea from the additive attention mechanism used by [7, 11] to calculate a relevancy preference score for each user, in hope to mitigate the *individual-level TRDI*. As shown in Figure 3 Phase 1, each item $c_i \in R^D$ in Clicked Item Sequence $Seq^{(c)} \in R^{L\times D}$ has a correspondence of its past trigger item $t_i \in R^D$ in Trigger Item Sequence $Seq^{(t)} \in R^{L\times D}$, where L is a fixed number of past clicks by the user and D is the embedding dimension. Since $Seq^{(c)}$ records the sequence of items clicked by a user in TIR scenario, we can always identify the trigger item associated with each click. These corresponding trigger items collectively form $Seq^{(t)}$. The embeddings of both items in the pair are used to compute the relevance vector $r_i \in R^D$, representing the user's degree of preference for trigger-relevancy or trigger-diversity at each time step. The relevance vectors of all pairs are then concatenated and transformed to a condensed vector $RelSignal \in R^H$ to represent the user's propensity to trigger-diversity, where H is a user-defined dimension size. The $RelSignal$ vector can be concatenated with other intermediate representations and signals computed by other modules of the model into a gating network such as MMoE [13], to enhance the model's ability in personalizing trigger-diversity propensity for each user. Formally, it is as follows in Equation 1 and 2, where the implementation is similar to the additive attention method used in the local activation of DIN [7], but omitting the weighted sum pooling, and pass the weights to a feed-forward network instead:

$$r_i = a_1(c_i, t_i) \tag{1}$$

$$RelSignal = a_2(\{r_1, r_2, ..., r_L\}) \tag{2}$$

where $a_1(\cdot)$ and $a_2(\cdot)$ are two feed-forward networks.

3.2.2 *Second Phase of DPSA*

In the second phase, we construct the Enhanced User Click Sequence $Seq^{(ce)} \in R^{L\times D}$ to enhance the model's understanding of user's differing propensity to trigger-diversity across categories of trigger items, thereby tackling the *trigger-item-level TRDI*. The entire second phase of DPSA can be abstracted as two independent attention operations $\mathrm{attn}_1(\cdot)$ and $\mathrm{attn}_2(\cdot)$ in equation 6 and 7.

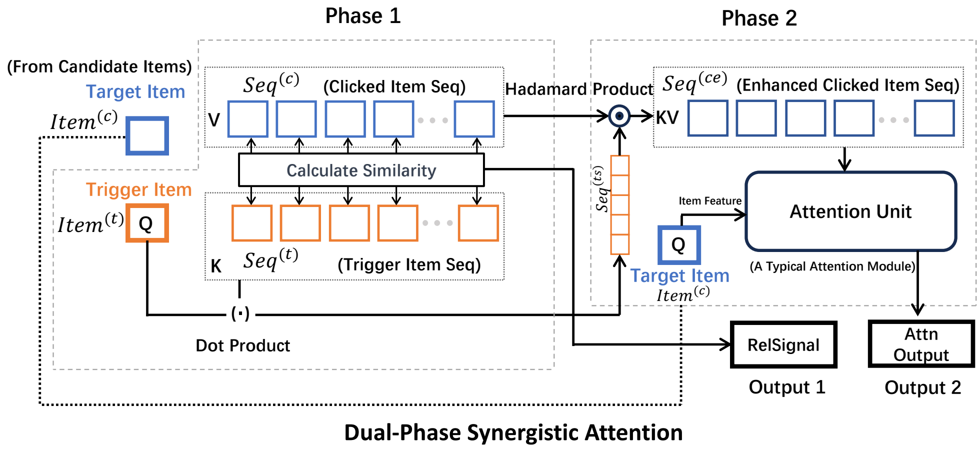

Figure 3: DPSA 1) In Phase 1, by calculating the relevance between $Seq^{(c)}$ and $Seq^{(t)}$, the user's Relevance Signal vector (RelSignal) is obtained. This vector potentially alleviates the TRDI issue at the individual-level. 2) Phase 2 can be considered to consist of two consecutive attention mechanisms. and the final output can address the TRDI problem at the trigger-item-level.

Firstly, we define the current trigger item as $t_0 \in R^D$. Then compute the relevance of t_0 to each t_i in $Seq^{(t)}$ as $e_i \in R^D$, giving rise to the relevance vector sequence $Seq^{(ts)} \in R^{L\times D}$. The calculation of e_i shares a similar idea in equation 1:

$$e_i = a_3(t_0, t_i) \tag{3}$$

$$Seq^{(ts)} = a_4(\{e_1, e_2, ..., e_L\}) \tag{4}$$

where $a_3(\cdot)$ and $a_4(\cdot)$ are two feed-forward neural networks.

The semantics of $Seq^{(ts)}$ is the similarity between the current trigger item t_0 and each past trigger item t_i that the user has viewed before clicking c_i. This can also be deemed as a type of attention score following [7]. Intuitively, if a user has seen a trigger item t_i very similar to the current trigger item t_0, and the user decided to click on the item c_i at that time, it suggests that the user may also have a higher likelihood to click on items that are related to c_i. So more attention should be paid to c_i, ensuring it contributes more to the user interest modeling. As such, the enhanced $Seq^{(c)}$, denoted as $Seq^{(ce)}$, can be described by the hadamard product of $Seq^{(c)}$ and $Seq^{(ts)}$ in equation 5:

$$Seq^{(ce)} = Seq^{(c)} \odot Seq^{(ts)} \tag{5}$$

Aforementioned equations 3-5 can be abstracted as $\mathrm{attn}_1(\cdot)$ in equation 6:

$$Seq^{(ce)} = \mathrm{attn}_1(Q = t_0, K = Seq^{(t)}, V = Seq^{(c)}) \tag{6}$$

Following, we process $Seq^{(ce)}$ and the candidate item $c_0 \in R^{L\times D}$ with scaled-dot product [14]. The computation is as follows in 7:

$$\mathrm{AttnOutput} = \mathrm{attn}_2(Q = c_0, K = Seq^{(ce)}, V = Seq^{(ce)}) \tag{7}$$

4 Experiments

We experimented with the proposed DPSA with the CTR ranking model at Shopee PDP YMAL PaidAds team. In offline settings, we conducted ablation studies to evaluate the effectiveness of each phase of DPSA. The best-performing offline model, the complete version of DPSA, is then selected for a one-week online A/B test to assess its production effectiveness over the baseline, where many business metrics that are hard to evaluate offline are assessed. After proofing its business effectiveness via online A/B test and deploying DPSA as the new production baseline, a 5% traffic holdout experiment is conducted for over a month to assess the long-term effectiveness of our proposed method.

Model	Offline AUC	Δ AUC
base	0.7378	0.00%
base+DPSA Phase 1	0.7395	0.23%
base+DPSA Phase 2	0.7392	0.19%
base+DPSA	**0.7403**	0.34%

Table 1: Offline Performance on Industrial Dataset

Model	Δ CTR	Δ CTCVR	Δ Rev	Δ GMV
base	-	-	-	-
base+DPSA	+1.12%	+2.10%	+1.78%	+3.32%

Table 2: Online A/B Test for 7-day in All Regional Markets. Involved billions of item impressions.

Model	Item Type	Δ Impression	Δ CTR
base	Trigger-Relevant	-	-
	Trigger-Diverse	-	-
base+DPSA	Trigger-Relevant	+3.37%	+4.61%
	Trigger-Diverse	-7.00%	+16.4%

Table 3: Online Post Analysis in one of the Regional Markets

4.1 Experimental Setup

Datasets. To the best of our knowledge, there is no publicly available dataset for the TIR task. Hence, we collected 30 days of anonymous user-system interaction logs from Shopee PDP YMAL for training. In addition, to construct the pair of Seq^c and Seq^t described in 3.2, each collected sample also contains a PDP item (i.e. trigger item), plus the real-time behavior sequences containing the respective user's past click actions in Shopee PDP YMAL. The constructed training dataset takes the impression-only samples as negative samples and the clicked samples as positive samples. After negative down-sampling at the rate of 20%, more than 3 billion training samples remained. Due to confidential policy, we are unable to share our internal dataset, but the community can follow our way to construct the pair of Seq^c and $Seq^{(t)}$, thereby producing a dataset to benchmark the effectiveness of DPSA in their use cases.

Baseline. Due to confidentiality policy, we cannot disclose the structure of our baseline. Nonetheless, DPSA is proposed as a plugin-like module, which can be easily integrated into any deep-based neural network model frameworks such as DIN [7], DeepFM [9], DMT [8], Wide&Deep [15] etc. Thus DPSA's effectiveness in improving a TIR recommendation system can still be expected to be fairly benchmarked over any baseline. In our offline and online experiments, the baseline used is the previous production CTR ranking model at Shopee PDP YMAL PaidAds, which is a deep-based neural network.

Offline Metrics. In *offline evaluation*, metrics like NDCG, MAP, AUC, or GAUC are commonly used for the CTR ranking phase. As the amount of training data used for Shopee PDP YMAL PaidAds CTR ranking model is very large, we used AUC for its computational simplicity, so that the training process is more cost-effective. To ensure fairness, the baseline is also retrained with the same dataset in offline comparison. In the later section, we report the offline AUC performance of the proposed methods over baseline.

Online Metrics In *online evaluation*, many business metrics need to be holistically assessed for the A/B test. In the Shopee PDP YMAL PaidAds team, we considered the key metrics of our three key stakeholders: users, sellers, and the platform. To ensure user experience, metrics like CTR and Click-Through&Conversion Rate (CTCVR) [16] are monitored; for the sellers, while CTR and CTCVR are important to sellers' efficiency in gaining returns from Shopee, metrics like GMV and ROI (return of investment) are also vital. For the platform, we care about Revenue from advertisement, as well as ROI for the sellers so that Shopee's advertising service would remain attractive to the sellers and stay sustainable in the long run.

4.2 Experimental Results

Offline. From Table 1, both Phase 1 and Phase 2 of DPSA improved AUC as an individual enhancement module. When both phases are integrated, the improvements from each phase remain largely orthogonal and give the best-performing model in the offline assessment, which is then used for the online A/B test. Although the percentage increase in AUC might seem small in our reporting in Table 1, a thousandth of a point increase in offline AUC is often a significant improvement in the real-world data of the recommendation advertising industry.

Online. The Shopee advertising system is a see-saw game involving three parties: users, sellers, and the platform. The see-saw effect in an advertising system is common, where improvement in one metric may come at the expense of the other. For example, when the platform decides to charge higher CPC (Cost per Click) or CPM (Cost per Mile) non-discriminatively, the revenue will surely increase in the short term, but sellers' GMV will not improve as the recommendation efficiency is not enhancec, thus the rise in platform revenue comes at the expense of sellers' ROI (Return of Investment). The advertising system could then enter a vicious cycle as sellers may abandon the advertising service due to cost-efficiency issues, and the platform will not be sustained in the long run. In our experiment, we trained a new CTR ranking model, which is only optimized for one of the key metrics - CTR. Therefore, to make an informed decision on whether to roll out a model to a production baseline, metrics from all three parties need to be considered holistically. In our online A/B test settings, 10% traffic is allocated for the control group, and the other 10% traffic for the treatment group. As PDP YMAL is the largest recommendation entrance in Shopee, each 10% traffic will give rise to billions of impressions over the 7-day A/B window. A significance test was also conducted to ensure the statistical significance of the A/B results. From Table 2, the 7-day online performance of DPSA significantly outperformed the baseline in all key dimensions for our key stakeholders. Notably in Table 2, as GMV improved much more than Revenue, the ROI for sellers is significantly improved, making Shopee's advertising service more efficient for the sellers and could benefit the growth and expansion of the platform in a virtuous cycle.

Post Analysis We also conducted post-analyses on the online performance of trigger-relevant and trigger-diverse recommendations. From Table 3, we present an analysis of one of Shopee's regional markets. The impressions of trigger-diverse recommendations were reduced by over 7%, and that of trigger-relevant recommendations increased by 4.61%. However, the CTR of both trigger-relevant and trigger-diverse recommendations improved. This implies that the efficiency in the distribution of the two types of recommendations has been enhanced with DPSA. Besides, the relative CTR improvement of trigger-diverse recommendations is much higher than that of trigger-relevant recommendations, showing that DPSA has improved the efficiency of traffic usage, particularly for trigger-diverse recommendations, thereby reducing traffic wastage.

5 Conclusions

In this paper, we focused on studying the TRDI issue in Shopee's largest recommendation ads scenario, PDP YMAL. Our contribution has two main components. Firstly, we propose a systematic approach to analyze the extent TRDI problem in TIR scenarios, breaking it down into two levels: Individual-level and Trigger-item-level. We hope that via this approach, the community will be able to assess the extent of TRDI in their business scenarios, and hopefully spark more targeted optimization efforts for TIR. Secondly, we proposed the Dual-Phase Synergistic Attention (DPSA) as a solution to alleviate the TRDI problem. DPSA comprises two stages and is uniquely designed with a two-phase attention mechanism that explicitly addresses the TRDI issue for each level. DPSA is designed to be a plugin-like module that can be easily integrated into any deep-based model architecture. Our final experimental results thoroughly validated the effectiveness of DPSA. Ultimately, DPSA was fully deployed in Shopee's YMAL advertisement's CTR prediction model across all regional markets, significantly optimizing various key metrics for all key stakeholders. In the future, DPSA, as a plugin-like module, will also be applied to multiple advertising recommendation stages, such as recall and pre-ranking, as well as in more TIR scenarios at Shopee. While TIR scenarios are prevalent in the industry, they have not yet received considerable attention from the research field and open-source community. A fundamental problem hindering the development of TIR research direction is the lack of a benchmark dataset for the community. Unfortunately, the company's confidential policy prohibits us from releasing our internal dataset. Still, we hope that as we share our method in constructing the pair of Clicked Item Sequence $Seq^{(c)}$ and Trigger Item Sequence $Seq^{(t)}$, others in the community will be able to publicize TIR-specific benchmark dataset for the community.

6 Acknowledgements

We would like to acknowledge the Shopee PaidAds Data team for constructing the trigger item into a user-behavior sequence, which greatly eases the algorithm exploration. We would like to acknowledge the Shopee PaidAds AI Platform team for optimizing the DPSA structure for inference so that the DPSA structure could run efficiently when serving online.

References

[1] Zihan Lin, Hui Wang, Jingshu Mao, Wayne Xin Zhao, Cheng Wang, Peng Jiang, and Ji-Rong Wen. Feature-aware diversified re-ranking with disentangled representations for relevant recommendation. In Aidong Zhang and Huzefa Rangwala, editors, *KDD '22: The 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, August 14 - 18, 2022*, pages 3327–3335. ACM, 2022.

[2] Qijie Shen, Hong Wen, Wanjie Tao, Jing Zhang, Fuyu Lv, Zulong Chen, and Zhao Li. Deep interest highlight network for click-through rate prediction in trigger-induced recommendation. In Frédérique Laforest, Raphaël Troncy, Elena Simperl, Deepak Agarwal, Aristides Gionis, Ivan Herman, and Lionel Médini, editors, *WWW '22: The ACM Web Conference 2022, Virtual Event, Lyon, France, April 25 - 29, 2022*, pages 422–430. ACM, 2022.

[3] Ruobing Xie, Rui Wang, Shaoliang Zhang, Zhihong Yang, Feng Xia, and Leyu Lin. Real-time relevant recommendation suggestion. In Liane Lewin-Eytan, David Carmel, Elad Yom-Tov, Eugene Agichtein, and Evgeniy Gabrilovich, editors, *WSDM '21, The Fourteenth ACM International Conference on Web Search and Data Mining, Virtual Event, Israel, March 8-12, 2021*, pages 112–120. ACM, 2021.

[4] Haolun Wu, Yansen Zhang, Chen Ma, Fuyuan Lyu, Fernando Diaz, and Xue Liu. A survey of diversification techniques in search and recommendation. *CoRR*, abs/2212.14464, 2022.

[5] Jennifer Gillenwater, Alex Kulesza, Emily B. Fox, and Benjamin Taskar. Expectation-maximization for learning determinantal point processes. In Zoubin Ghahramani, Max Welling, Corinna Cortes, Neil D. Lawrence, and Kilian Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 3149–3157, 2014.

[6] Shuai Yang, Lixin Zhang, Feng Xia, and Leyu Lin. Graph exploration matters: Improving both individual-level and system-level diversity in wechat feed recommender. *CoRR*, abs/2306.00009, 2023.

[7] Guorui Zhou, Xiaoqiang Zhu, Chengru Song, Ying Fan, Han Zhu, Xiao Ma, Yanghui Yan, Junqi Jin, Han Li, and Kun Gai. Deep interest network for click-through rate prediction. In Yike Guo and Faisal Farooq, editors, *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018*, pages 1059–1068. ACM, 2018.

[8] Yulong Gu, Zhuoye Ding, Shuaiqiang Wang, Lixin Zou, Yiding Liu, and Dawei Yin. Deep multifaceted transformers for multi-objective ranking in large-scale e-commerce recommender systems. In Mathieu d'Aquin, Stefan Dietze, Claudia Hauff, Edward Curry, and Philippe Cudré-Mauroux, editors, *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020*, pages 2493–2500. ACM, 2020.

[9] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. Deepfm: A factorization-machine based neural network for CTR prediction. In Carles Sierra, editor, *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, pages 1725–1731. ijcai.org, 2017.

[10] Qi Pi, Guorui Zhou, Yujing Zhang, Zhe Wang, Lejian Ren, Ying Fan, Xiaoqiang Zhu, and Kun Gai. Search-based user interest modeling with lifelong sequential behavior data for click-through rate prediction. In Mathieu d'Aquin, Stefan Dietze, Claudia Hauff, Edward Curry, and Philippe Cudré-Mauroux, editors, *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020*, pages 2685–2692. ACM, 2020.

[11] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

[12] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 5998–6008, 2017.

[13] Jiaqi Ma, Zhe Zhao, Xinyang Yi, Jilin Chen, Lichan Hong, and Ed H. Chi. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In Yike Guo and Faisal Farooq, editors, *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018*, pages 1930–1939. ACM, 2018.

[14] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 5998–6008, 2017.

[15] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, Rohan Anil, Zakaria Haque, Lichan Hong, Vihan Jain, Xiaobing Liu, and Hemal Shah. Wide & deep learning for recommender systems. In Alexandros Karatzoglou, Balázs Hidasi, Domonkos Tikk, Oren Sar Shalom, Haggai Roitman, Bracha Shapira, and Lior Rokach, editors, *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, DLRS@RecSys 2016, Boston, MA, USA, September 15, 2016*, pages 7–10. ACM, 2016.

[16] Xiao Ma, Liqin Zhao, Guan Huang, Zhi Wang, Zelin Hu, Xiaoqiang Zhu, and Kun Gai. Entire space multi-task model: An effective approach for estimating post-click conversion rate. In Kevyn Collins-Thompson, Qiaozhu Mei, Brian D. Davison, Yiqun Liu, and Emine Yilmaz, editors, *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR 2018, Ann Arbor, MI, USA, July 08-12, 2018*, pages 1137–1140. ACM, 2018.