# Modeling Temporal Positive and Negative Excitation for Sequential Recommendation

Chengkai Huang*
chengkai.huang1@unsw.edu.au
The University of New South Wales
Sydney, NSW, Australia

Shoujin Wang
shoujin.wang@uts.edu.au
University of Technology Sydney
Sydney, NSW, Australia

Xianzhi Wang
XIANZHI.WANG@uts.edu.au
University of Technology Sydney
Sydney, NSW, Australia

Lina Yao
lina.yao@unsw.edu.au
CSIRO's Data 61 and UNSW
Sydney, NSW, Australia

SIGIR 2023

**Reported by Nengqiang Xiang**

# Introduction

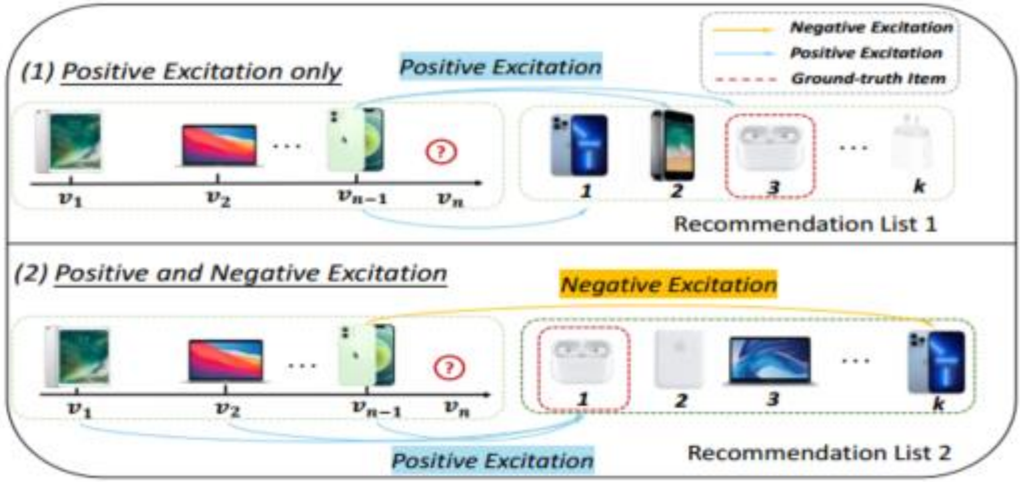
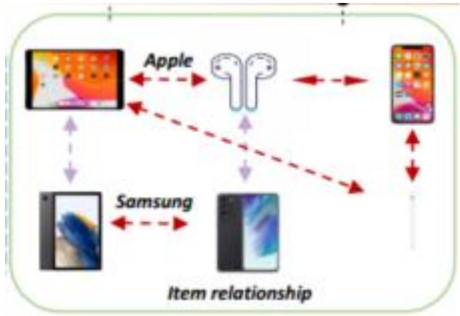



**Figure 1: An example of recommendations via modeling positive excitation only (existing methods) and modeling both positive and negative excitation (our proposal). Clearly, the latter achieves better performance via ranking the ground-truth next item AirPods at the Top-1 position in the recommendation list.**

Overlooking users' static interest revealed by some static attribute information of items, (category, brand).

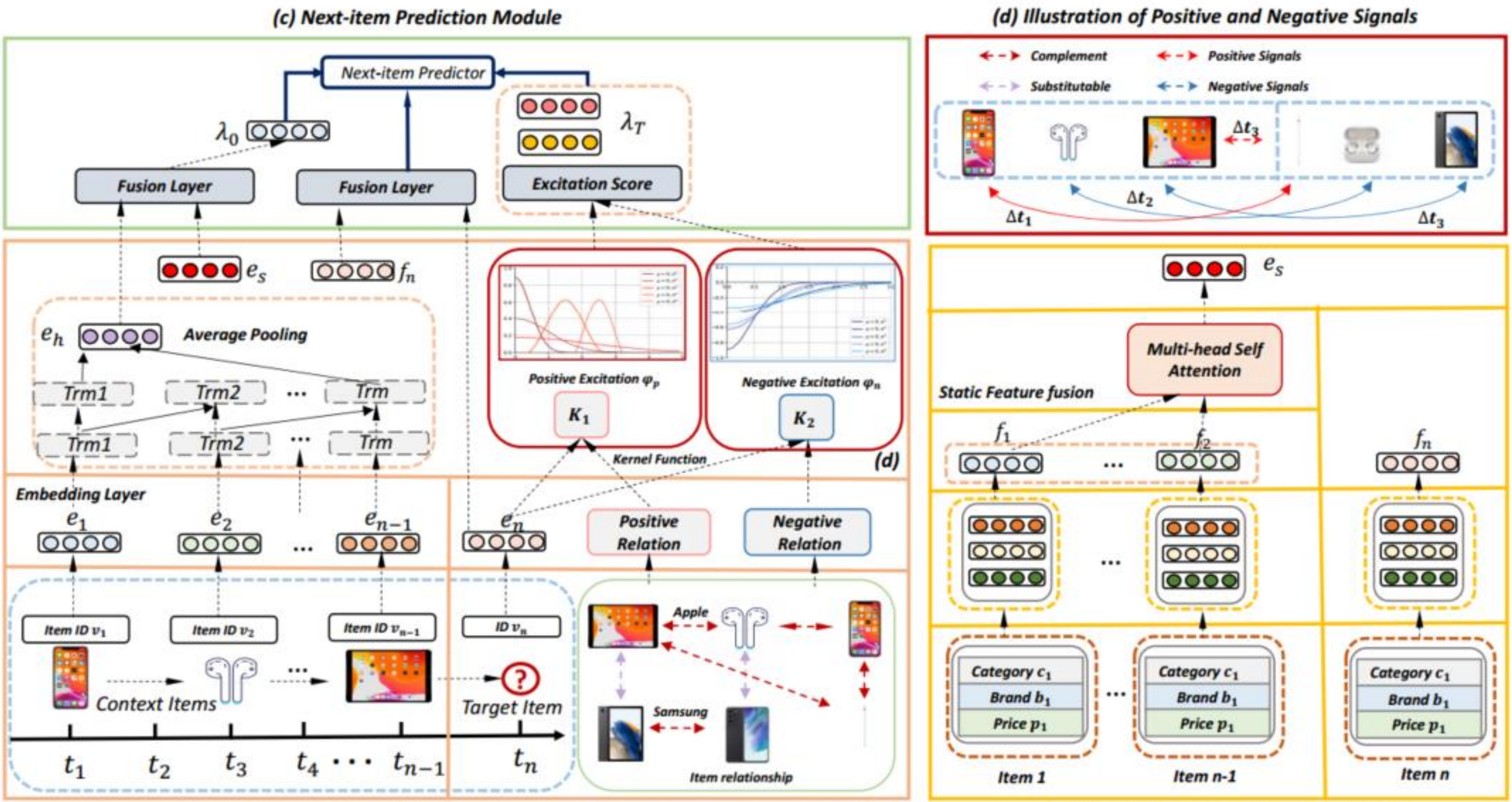
Existing works often only consider the positive excitation of a user's historical interactions.

In this paper, the author proposed modeling both static interest and negative excitation for dynamic interest.



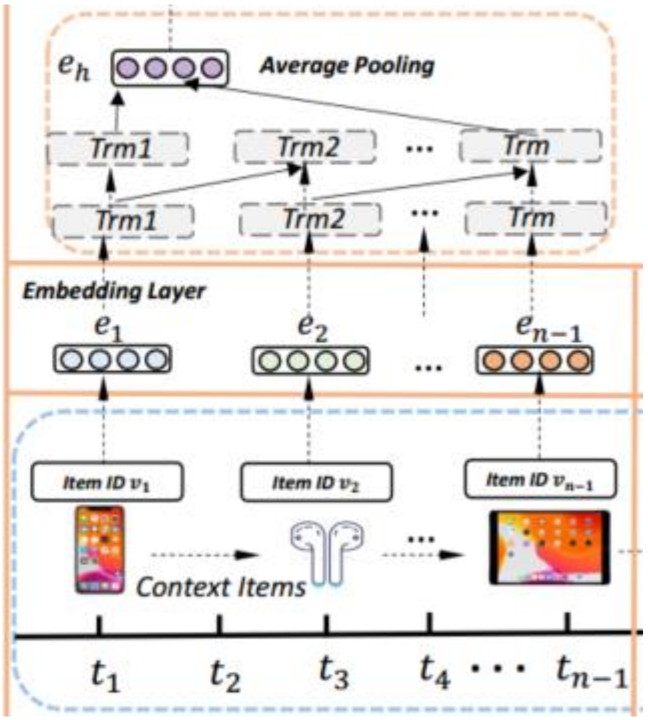

complementary
also_buy ($r_1$)
share_brand ($r_3$)

substitute
also_view ($r_2$)
similar_item ($r_4$)




(a) User Dynamic Interest Modelling (DIM) with Temporal Positive and Negative Excitation (TPNE)

(b) User Static Interest Modelling (SIM)




## PRELIMINARIES:

$$\mathcal{U} = \{u_1, u_2, ..., u_{|U|}\}$$

$$\mathcal{V} = \{v_1, v_2, ..., v_{|V|}\}$$

$$H_u = \{v_1, v_2, ..., v_n\}$$

category $c_i$

brand $b_i$

price $p_i$

interaction timestamp $t_i$

item embedding matrix $E \in R^{|V| x d}$

$$e_v = E(v) \in R^{1xd}$$

$u$'s item embedding $E(H_u)$

## Hawkes Processes in Sequential Modeling

$$\lambda(t) = \mu(t) + \sum_{t_i<t} \varphi(t - t_i), \tag{1}$$

## Users' Dynamic Interest Modeling

$$\lambda_T(t) = \lambda_0 + \sum_{i:t_i<t_n} \varphi(t_n - t_i), \tag{2}$$

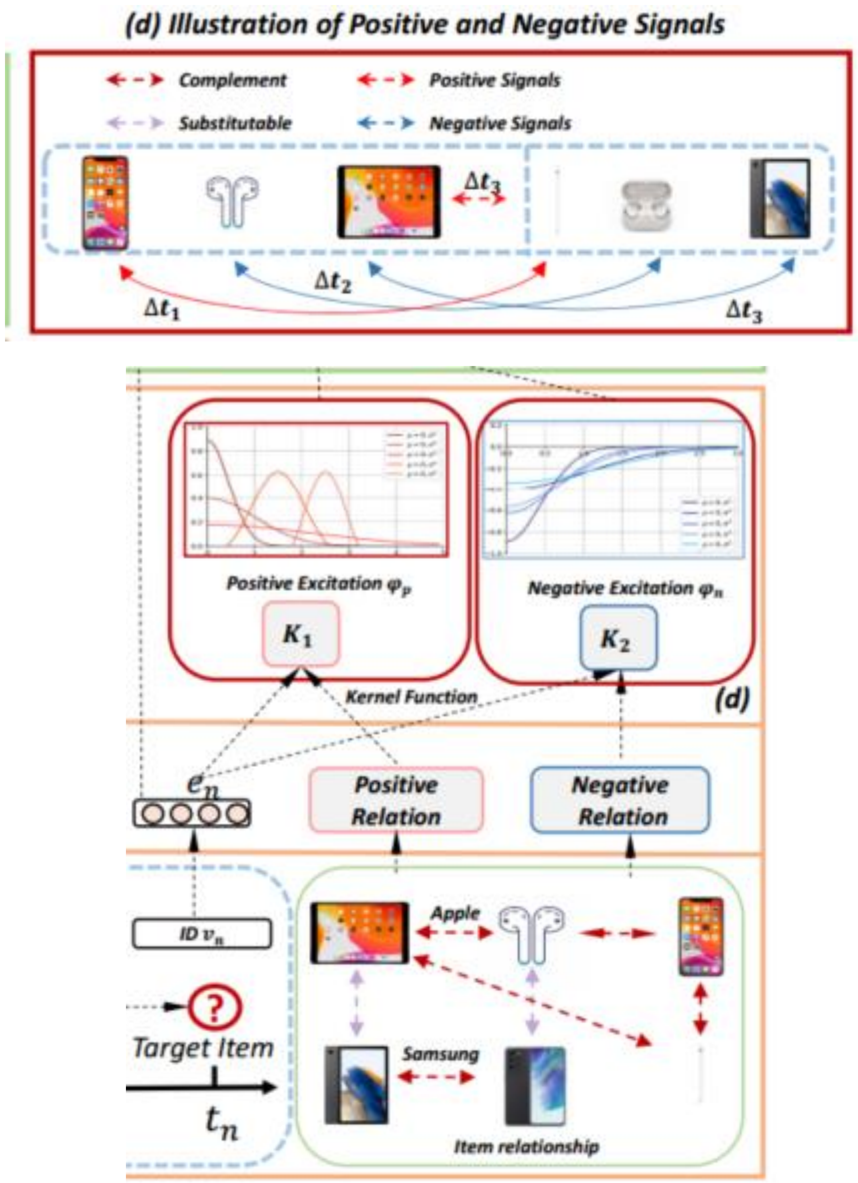
$$\lambda_0 = e_h^T e_{v_n} + u_b + i_b, \tag{3}$$

$$A_i = Att((E(H_u)+POS_i)W^Q, (E(H_u)+POS_i)W^K, (E(H_u)+POS_i)W^V), \tag{4}$$

$$Att_i = Attention(Q, K, V) = \mathrm{softmax}\left(\frac{QK^\top}{\sqrt{d}}\right)V, \tag{5}$$

$$E(H_u) = LayerNorm(H_u + Dropout(FFN(A_i))). \tag{6}$$

$$e_h = \frac{1}{|N| - 1} \sum_{i=1}^{n-1} E(H_u)_i, \tag{7}$$




## Positive Excitation Learning and Negative Excitation Learning

$$\lambda_T(t) = \lambda_0 + \sum_{i:t_i<t_n} \varphi_p(t_n - t_i) - \sum_{j:t_j<t_n} \varphi_n(t_n - t_j), \quad (8)$$

## Positive Excitation Learning

$$\varphi_p(t_n - t_i) = \sum_{i:t_i<t} I_{rp}(v_i, v_n)\mathcal{K}_1(t_n - t_i), \quad (9)$$

$$\mathcal{K}_1^i(\Delta t_1) = N\left(\Delta t_1 \mid 0, \sigma_1^v\right) + N\left(\Delta t_1 \mid \mu_2^v, \sigma_2^v\right). \quad (10)$$

$r \in \{r1, r2, r3, r4\}$

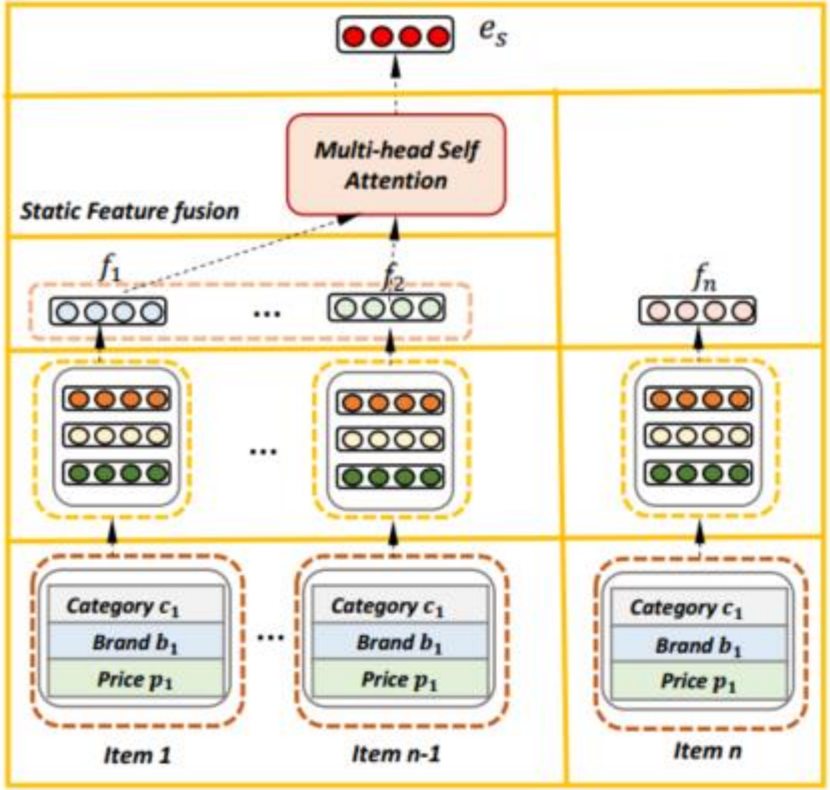
complementary
- also_buy ($r_1$)
- share_brand ($r_3$)

## Negative Excitation Learning

substitute
- also_view ($r_2$)
- similar_item ($r_4$)

$$\varphi_n(t_n - t_j) = \sum_{j:t_j<t_n} I_{rn}(v_j, v_n)\mathcal{K}_2(t_n - t_j), \quad (11)$$

$$\mathcal{K}_2(\Delta t_2) = -N\left(\Delta t_2 \mid 0, \sigma_3^v\right), \quad (12)$$



(b) User Static Interest Modelling (SIM)

## Users' Static Interest Modeling

category embedding matrix $C \in R^{|C| \times d}$

brand embedding matrix $B \in R^{|B| \times d}$

price embedding matrix $P \in R^{|P| \times d}$

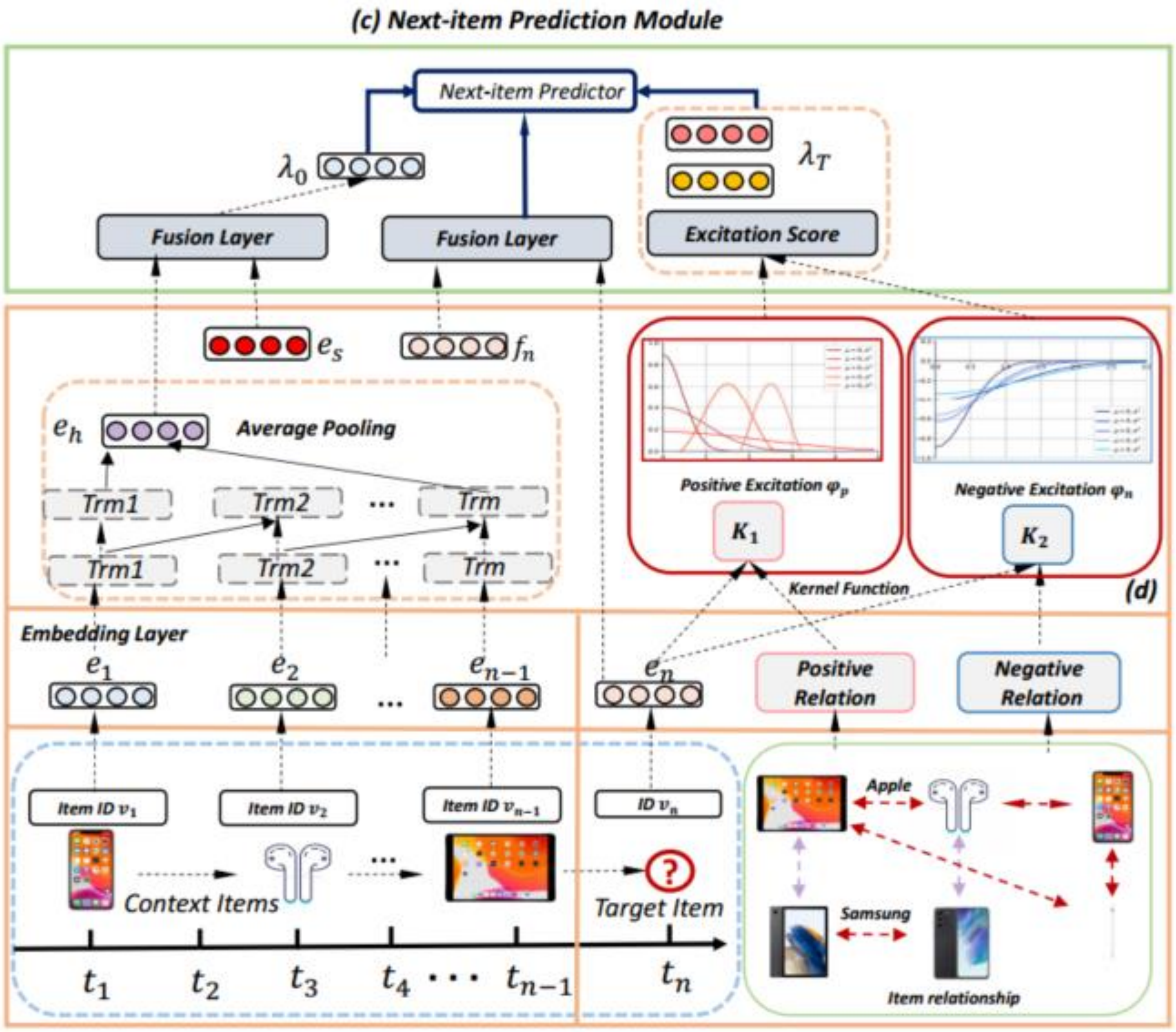
$$f_i = c_i + b_i + p_i, \quad (13)$$

$$H_f = Att(FW^Q, FW^K, FW^V). \quad (14)$$

$$M_f = Multihead(F) = Concat(h_1, h_2, ..., h_{l_f})W^O, \quad (15)$$

$$h_i = Att(FW_i^Q, FW_i^K, FW_i^V), \quad (16)$$

$$e_s = \frac{1}{|N| - 1} \sum_{i=1}^{|N|-1} H_f, \quad (19)$$




(a) User Dynamic Interest Modelling (DIM) with Temporal Positive and Negative Excitation (TPNE)

## Next-item Prediction

$$g = \sigma(W_1 e_s + W_2 e_h + b), \tag{20}$$

$$e_f = g \odot e_s + (1 - g) \odot e_h, \tag{21}$$

$$\mathcal{L}_r = -\sum_{u \in \mathcal{U}} \sum_{i=1}^{N_u} \log \sigma \left( \hat{y}_{ui} - \hat{y}_{uj} \right), \tag{22}$$

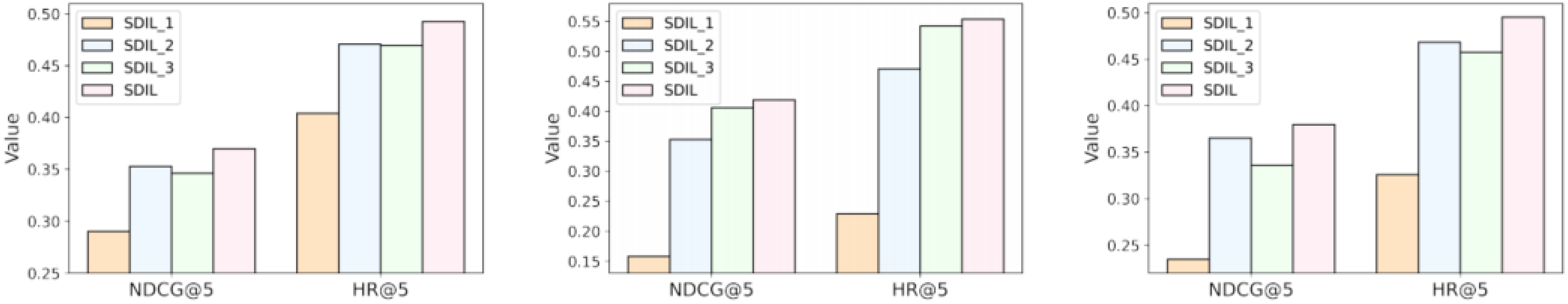
$$\hat{y}_{ui} = e_f^T e_i + \lambda_{T,i}, \quad \hat{y}_{uj} = e_f^T e_j + \lambda_{T,j} \tag{23}$$

**Table 2: Overall performance. Bold scores represent the highest results of all methods. Underlined scores stand for the second-highest results. Our model achieves the state-of-the-art result among all baseline models. * means the improvement is significant at $p < 0.05$.**

| Dataset | Metric | BPR | GRU4Rec | Caser | NARM | SASRec | TiSASRec | SLRS+ | Chorus | AHMP | KDA | SDIL | Improv. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Beauty | HR@5 | 0.3317 | 0.3202 | 0.3210 | 0.3334 | 0.4004 | 0.3872 | 0.4339 | 0.4536 | 0.4566 | <u>0.4860</u> | **0.4926*** | 1.36% |
| | HR@10 | 0.4355 | 0.4311 | 0.4345 | 0.4462 | 0.5074 | 0.4980 | 0.5337 | 0.5698 | 0.5519 | <u>0.5997</u> | **0.6128*** | 2.18% |
| | HR@20 | 0.5505 | 0.5693 | 0.5757 | 0.5823 | 0.6268 | 0.6179 | 0.6361 | 0.6838 | 0.6599 | <u>0.7144</u> | **0.7323*** | 2.51% |
| | NDCG@5 | 0.2361 | 0.2271 | 0.2246 | 0.2348 | 0.2923 | 0.2904 | 0.3319 | 0.3386 | 0.3496 | <u>0.3648</u> | **0.3698*** | 1.37% |
| | NDCG@10 | 0.2697 | 0.2628 | 0.2612 | 0.2712 | 0.3268 | 0.3181 | 0.3642 | 0.3762 | 0.3803 | <u>0.4016</u> | **0.4088*** | 1.79% |
| | NDCG@20 | 0.2987 | 0.2976 | 0.2967 | 0.3055 | 0.3569 | 0.3483 | 0.3900 | 0.4050 | 0.4076 | <u>0.4306</u> | **0.4390*** | 1.95% |
| | MRR | 0.2363 | 0.2271 | 0.2246 | 0.2366 | 0.2923 | 0.2904 | 0.3319 | 0.3386 | 0.3421 | <u>0.3549</u> | **0.3610*** | 1.72% |
| Cellphone | HR@5 | 0.3387 | 0.3015 | 0.3937 | 0.4168 | 0.4586 | 0.4520 | 0.4696 | 0.4697 | 0.5045 | <u>0.5497</u> | **0.5538*** | 0.75% |
| | HR@10 | 0.4528 | 0.4301 | 0.5309 | 0.5509 | 0.5810 | 0.5767 | 0.5641 | 0.5929 | 0.6132 | <u>0.6745</u> | **0.6792*** | 0.70% |
| | HR@20 | 0.5852 | 0.5918 | 0.6810 | 0.6974 | 0.7067 | 0.7022 | 0.6637 | 0.7152 | 0.7284 | <u>0.7923</u> | **0.8028*** | 1.33% |
| | NDCG@5 | 0.2430 | 0.2085 | 0.2800 | 0.2995 | 0.3412 | 0.3344 | 0.3634 | 0.3530 | 0.3852 | <u>0.4119</u> | **0.4188*** | 1.69% |
| | NDCG@10 | 0.2798 | 0.2498 | 0.3243 | 0.3429 | 0.3809 | 0.3748 | 0.3939 | 0.3929 | 0.4204 | <u>0.4523</u> | **0.4595*** | 1.59% |
| | NDCG@20 | 0.3131 | 0.2905 | 0.3622 | 0.3799 | 0.4126 | 0.4065 | 0.4191 | 0.4238 | 0.4495 | <u>0.4821</u> | **0.4908*** | 1.80% |
| | MRR | 0.2453 | 0.2271 | 0.2246 | 0.2969 | 0.2923 | 0.2904 | 0.3319 | 0.3386 | 0.3747 | <u>0.3666</u> | **0.4049*** | 10.45% |
| Toys | HR@5 | 0.2897 | 0.2902 | 0.2898 | 0.3173 | 0.3687 | 0.3475 | 0.4368 | 0.4124 | 0.4603 | <u>0.4805</u> | **0.4953*** | 3.08% |
| | HR@10 | 0.3897 | 0.4060 | 0.4103 | 0.4336 | 0.4767 | 0.4608 | 0.5345 | 0.5203 | 0.5587 | <u>0.5882</u> | **0.6069*** | 3.18% |
| | HR@20 | 0.5061 | 0.5546 | 0.5590 | 0.5777 | 0.6018 | 0.6003 | 0.6440 | 0.6443 | 0.6621 | <u>0.7019</u> | **0.7248*** | 3.26% |
| | NDCG@5 | 0.2068 | 0.1974 | 0.1947 | 0.2206 | 0.3023 | 0.2535 | 0.3490 | 0.3132 | 0.3600 | <u>0.3660</u> | **0.3797*** | 3.74% |
| | NDCG@10 | 0.2390 | 0.2348 | 0.2336 | 0.2581 | 0.3140 | 0.2901 | 0.3804 | 0.3480 | 0.3918 | <u>0.4007</u> | **0.4157*** | 3.74% |
| | NDCG@20 | 0.2683 | 0.2721 | 0.2710 | 0.2944 | 0.3339 | 0.3253 | 0.4081 | 0.3793 | 0.4179 | <u>0.4294</u> | **0.4454*** | 3.73% |
| | MRR | 0.2116 | 0.2271 | 0.2246 | 0.2244 | 0.2923 | 0.2904 | 0.3319 | 0.3386 | 0.3547 | <u>0.3666</u> | **0.3713*** | 1.28% |

# Experiments






(a) Beauty Dataset. (b) Cellphones Dataset. (c) Toys Dataset.

Figure 3: Ablation Study on the Model Performance. (HR@5 and NDCG@5) on different datasets.

**Table 3: Performance comparison between SDIL-TPE and SDIL. * means the improvement is significant at $p < 0.05$.**

| Dataset | Metrics | SDIL-TPE | **SDIL** |
|---|---|---|---|
| Beauty | *HR@5* | 0.4825 | 0.4926* |
| | *HR@10* | 0.6054 | 0.6128* |
| | *NDCG@5* | 0.3487 | 0.3698* |
| | *NDCG@10* | 0.4014 | 0.4088* |
| | *MRR* | 0.3534 | 0.3602* |
| Cellphones | *HR@5* | 0.5521 | 0.5538* |
| | *HR@10* | 0.6772 | 0.6792* |
| | *NDCG@5* | 0.4102 | 0.4188* |
| | *NDCG@10* | 0.4422 | 0.4595* |
| | *MRR* | 0.4038 | 0.4049* |
| Toys | *HR@5* | 0.4871 | 0.4953* |
| | *HR@10* | 0.5979 | 0.6069* |
| | *NDCG@5* | 0.3741 | 0.3797* |
| | *NDCG@10* | 0.3670 | 0.4157* |
| | *MRR* | 0.3670 | 0.3713* |

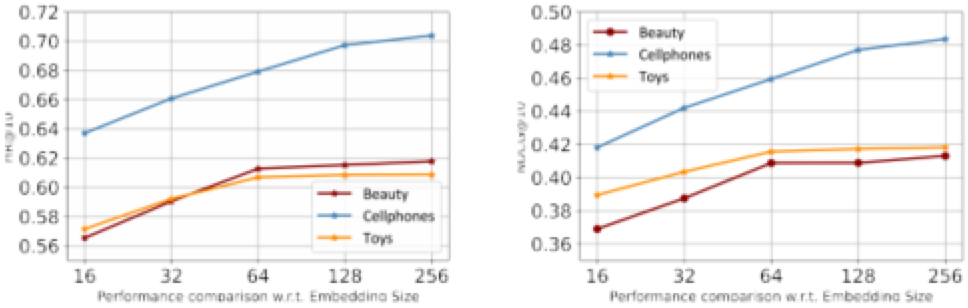



**Figure 4: Embedding size setting's effect on the model performance. (HR@5 and NDCG@10).**

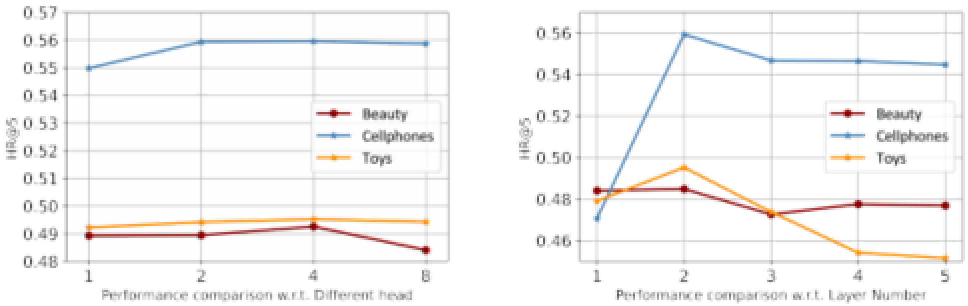



**Figure 5: Different transformer layers setting's effect on the model performance. (HR@10 and NDCG@10).**

# Experiments

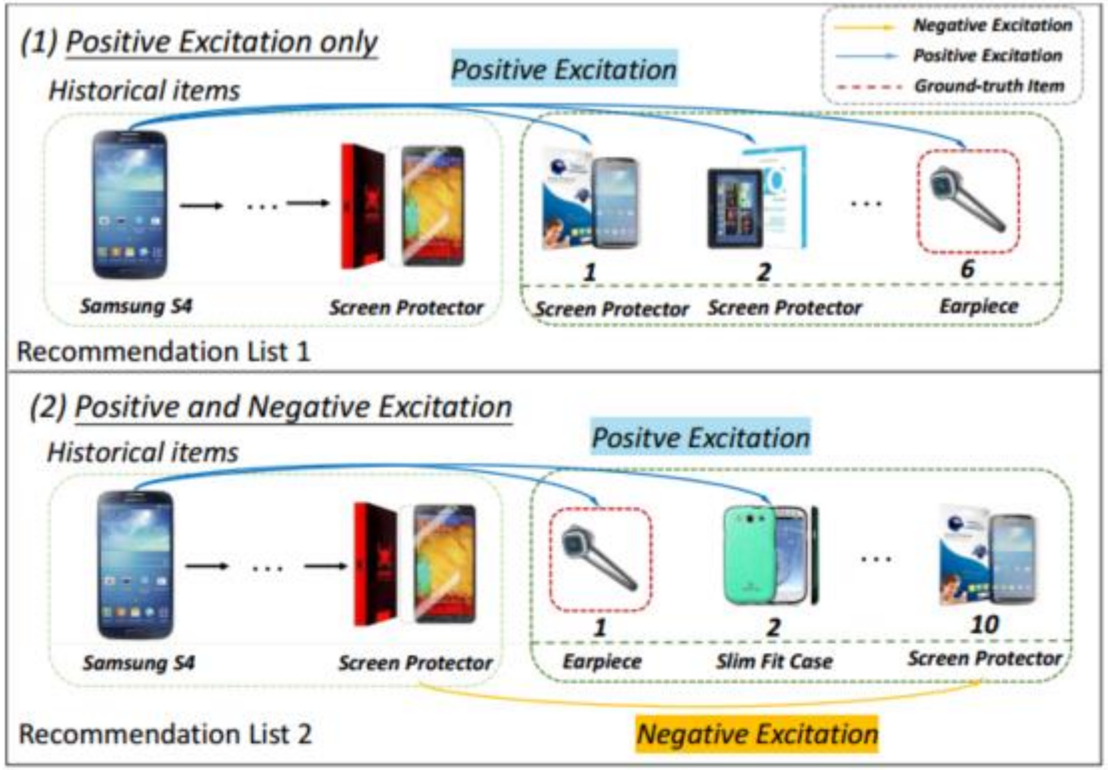



**Figure 6: Illustration of the ranking results of TPE and TPNE. The item highlighted in the red boxes is the ground-truth item.**

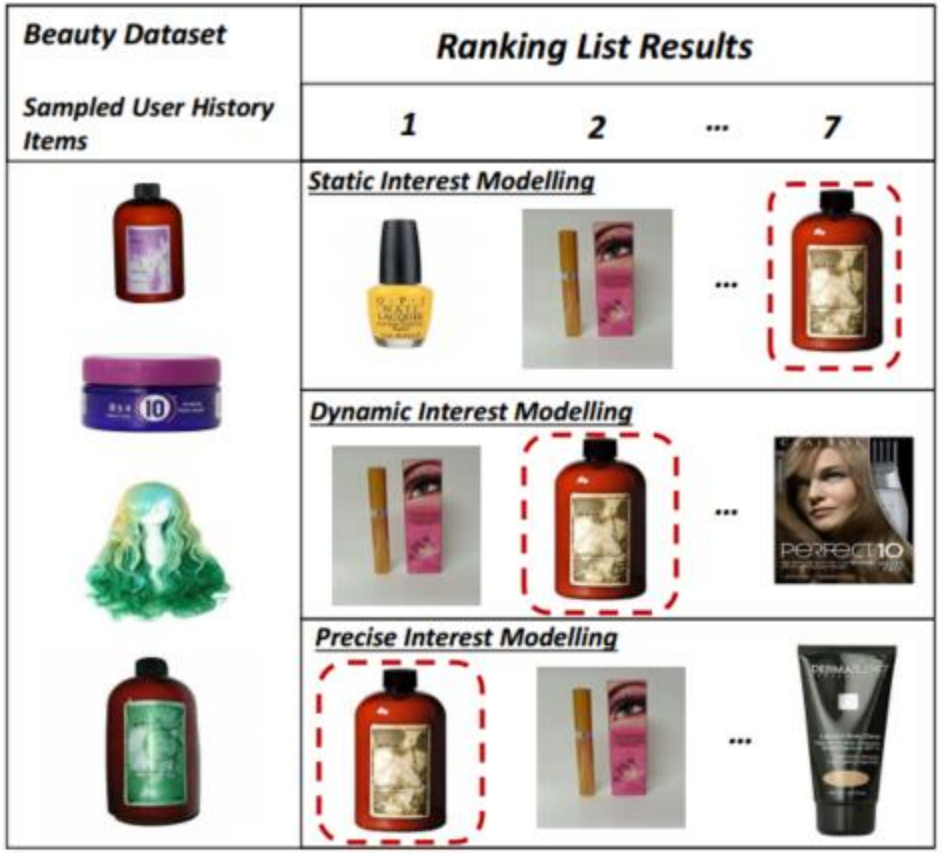



**Figure 7: Illustration of the ranking results of SIM, DIM and DSIM. The item highlighted in the red boxes is the ground-truth item.**