Residual Multi-Task Learner for Applied Ranking

Based on joint work with Cong Fu et al. Deployed in Shopee Search.

Zhiming Zhou | SUFE | 2024/12/07

Contents

01 Problem

02 Existing Methods

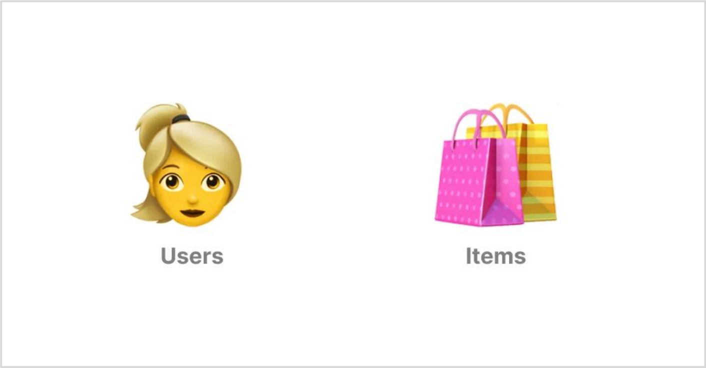
03 ResFlow

04 Online Deployment

01 Problem

Applied Ranking

Predicting which items users will be more interested in,



(e.g., search, recommendation)

based on user and item features and their interactions.

John	5	1	3	5
Tom	?	?	?	2
Alice	4	?	3	?

Multi-Task in Applied Ranking

When ranking items,

we may **estimate**

the click-through rate (CTR)

the add-to-cart rate (ATCR)

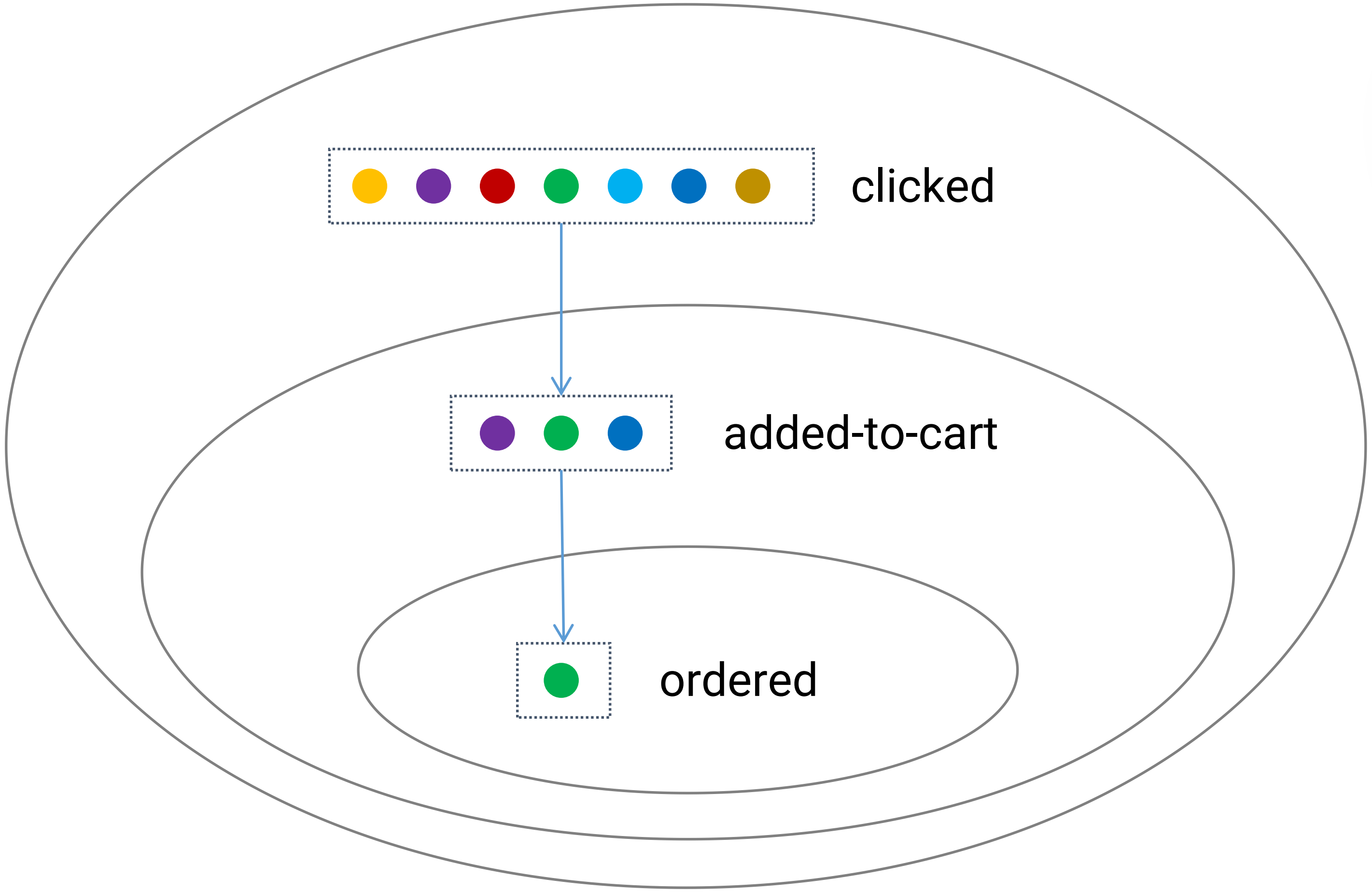
the final conversion rate (CTCVR)

Multi-Task
(same input, different but related outputs)

of a user w.r.t. items,

and jointly consider all these predictions.

Attribute 0 : Sparse Data



latter task, e.g., CTCVR estimation, has significantly sparser data → multi-task learning

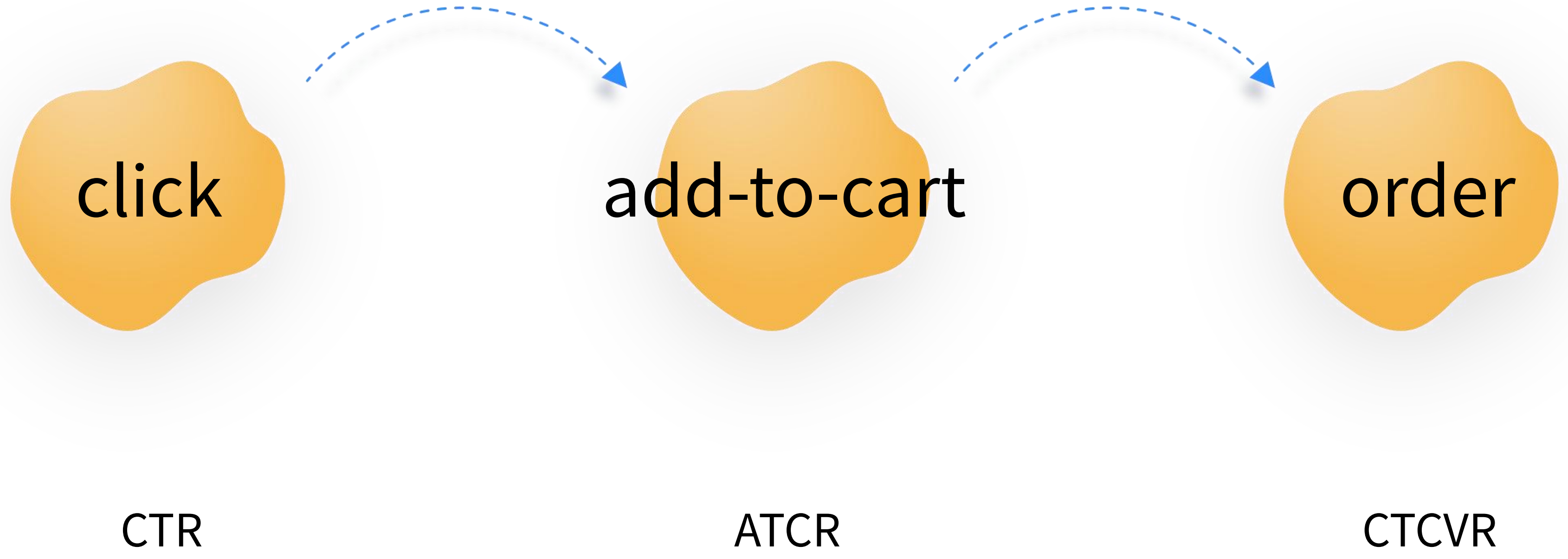
Attribute 1 : Large Scale & Real-Time

Practical application scenario:

- enormous items
- need to response in a second

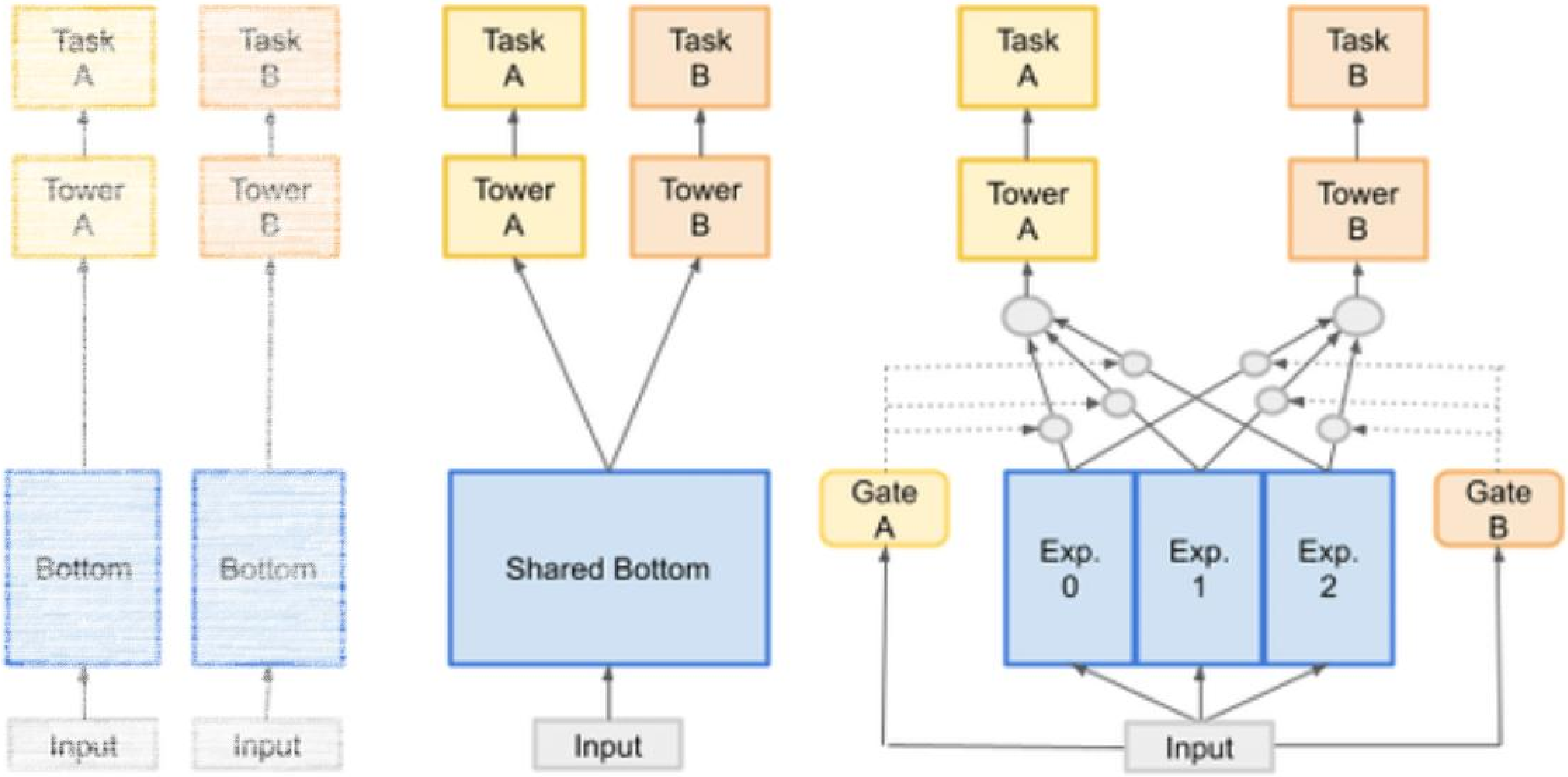
Ranking algorithm **esquires high efficiency**.

Attribute 2 : Sequential Dependency



02 Existing Methods

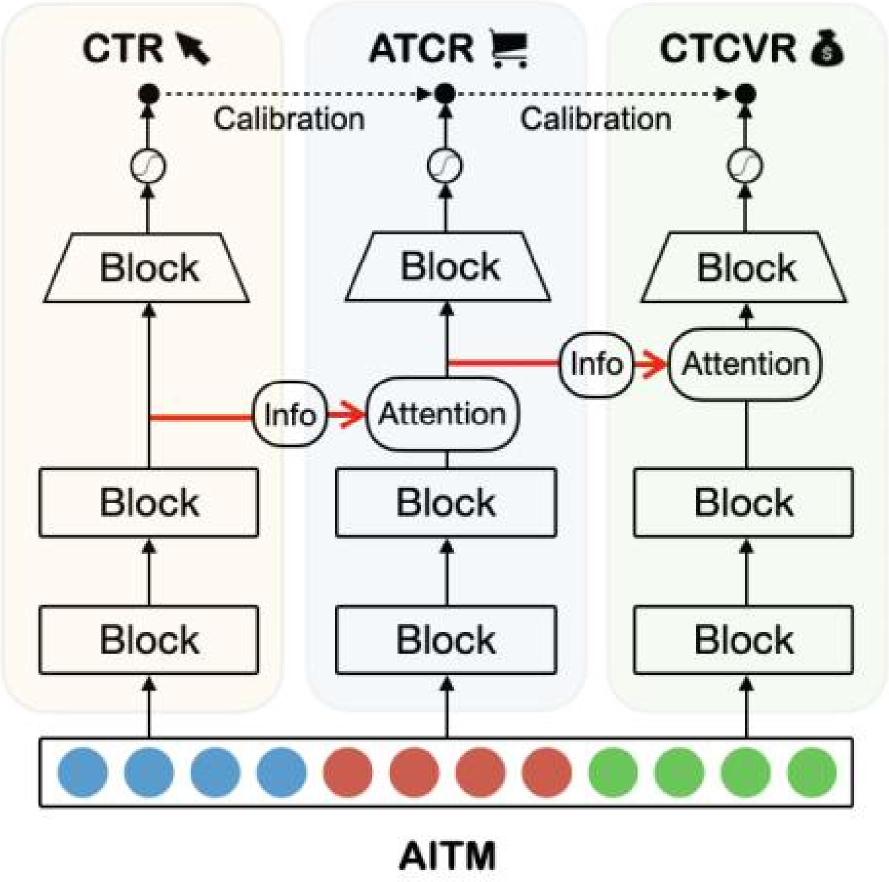
The Shared Bottom Class



(a) (b) (c)

indirect information transfer → inferior performance

Adaptive Information Transfer

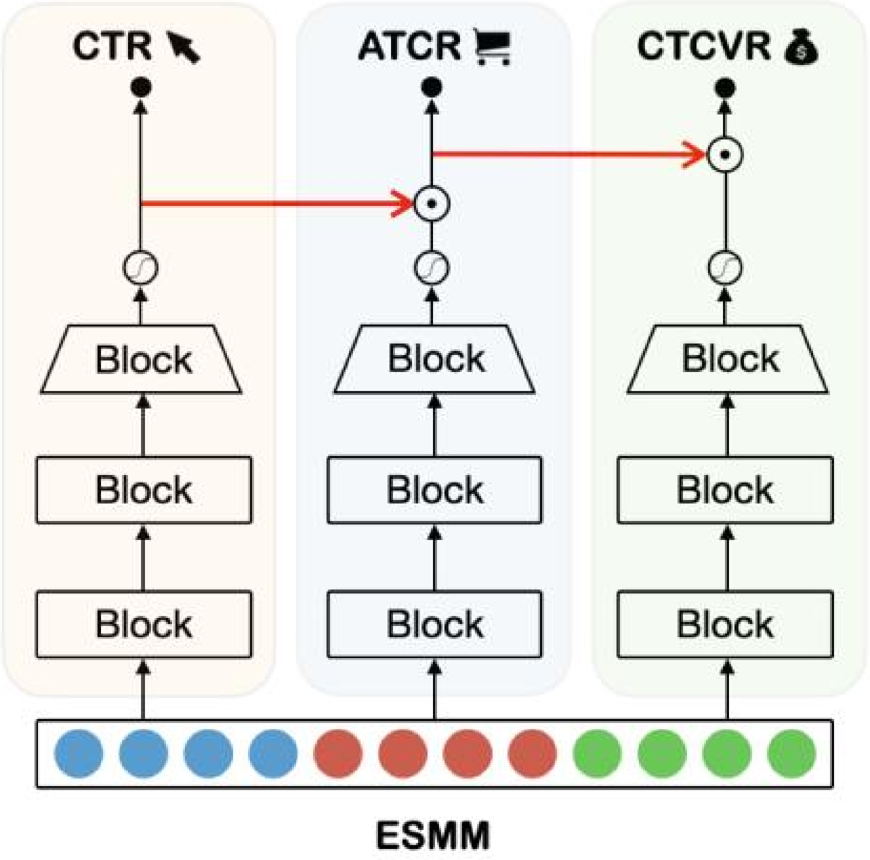


Transferring the last layer feature:

- via an attention-based module

- ☐ Computationally inefficient
- ☐ Inadequate information transfer

Entire Space Multitask Model (ESMM)



Transferring the output probability:

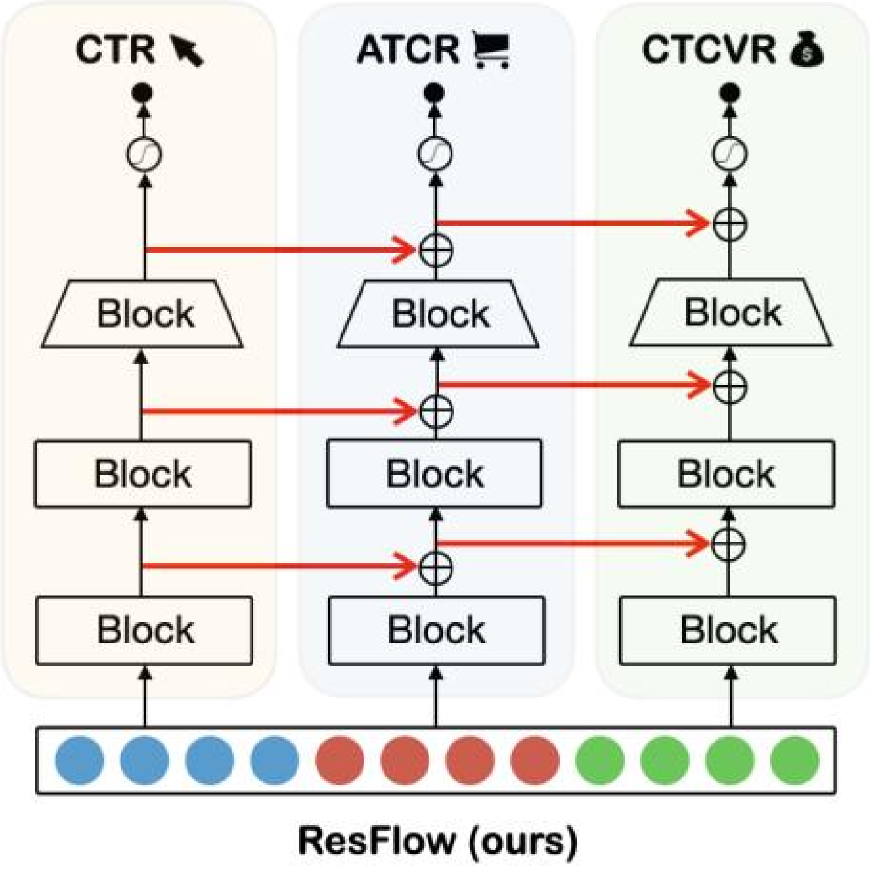
- multiplicatively
- modeling the conditional probability

☐ Light-weight

☐ Inadequate information transfer

03 ResFlow

Residual Multi-Task Learner (ResFlow)



Transferring <u>all features and the final logit</u>

- additively
- via cross-task residual connection

☐ light weight

☐ sufficient & effective information transfer

Results

Table 4: The AUC results of the CTCVR estimation task on offline e-commerce datasets. The best results are presented in bold font, while the second bests are marked with underlines. ResFlow consistently achieves the best.

Dataset	NSE	PLE	AITM	ESMM	ESCM^2-IPW	ESCM^2-DR	DCMT	ResFlow (ours)
S0	0.619±0.001	0.630±0.005	0.636 ± 0.002	0.632 ± 0.001	0.624 ± 0.007	0.622±0.006	<u>0.640 ± 0.001</u>	**0.656 ± 0.001**
S1	0.621±0.002	0.624±0.003	0.623 ± 0.001	0.623 ± 0.003	0.624 ± 0.002	0.623±0.003	<u>0.632 ± 0.002</u>	**0.647 ± 0.002**
S0&S1	0.623±0.002	0.635±0.001	0.637 ± 0.001	0.637 ± 0.003	0.623 ± 0.005	0.623±0.001	<u>0.634 ± 0.002</u>	**0.661 ± 0.001**
AliCCP	0.624±0.002	0.640±0.002	<u>0.644 ± 0.001</u>	0.641 ± 0.003	0.625 ± 0.005	0.627±0.003	0.643 ± 0.002	**0.664 ± 0.002**
AE-ES	0.861±0.002	0.873±0.002	0.876 ± 0.001	0.867 ± 0.003	0.868 ± 0.003	0.871±0.002	<u>0.886 ± 0.003</u>	**0.893 ± 0.001**
AE-FR	0.842±0.002	0.852±0.002	0.856 ± 0.002	0.851 ± 0.001	0.873 ± 0.007	0.870±0.005	<u>0.874 ± 0.003</u>	**0.885 ± 0.001**
AE-NL	0.831±0.003	0.844±0.002	0.843 ± 0.001	0.840 ± 0.002	0.854 ± 0.006	0.854±0.004	<u>0.858 ± 0.001</u>	**0.864 ± 0.001**
AE-US	0.826±0.001	0.851±0.003	0.843 ± 0.002	0.827 ± 0.002	0.843 ± 0.006	0.841±0.003	<u>0.863 ± 0.003</u>	**0.872 ± 0.001**
AE-RU	0.870±0.003	0.886±0.001	0.882 ± 0.002	0.878 ± 0.002	0.882 ± 0.006	0.880±0.006	<u>0.887 ± 0.001</u>	**0.913 ± 0.002**
Shopee-2	0.865±0.001	0.860±0.003	0.855 ± 0.006	<u>0.882 ± 0.001</u>	0.840 ± 0.012	0.812±0.041	0.867 ± 0.001	**0.902 ± 0.001**
Shopee-3	0.877±0.002	0.877±0.003	0.871 ± 0.003	<u>0.893 ± 0.001</u>	/	/	/	**0.910 ± 0.002**

Ablations

Table 5: Ablation results of ResFlow in terms of CTCVR AUC.

Model	AE-RU	Shopee-2
NSE	0.869 ± 0.003	0.865 ± 0.001
NSE + Feature Residual (FR)	0.897 ± 0.001	0.891 ± 0.001
NSE + FR (H1-only)	0.880 ± 0.002	0.868 ± 0.002
NSE + FR (H2-only)	0.898 ± 0.002	0.887 ± 0.001
NSE + Logit Residual (LR)	0.907 ± 0.002	0.896 ± 0.001
ResFlow (NSE + FR + LR)	$\mathbf{0.913 \pm 0.002}$	$\mathbf{0.902 \pm 0.001}$
ESMM	0.878 ± 0.002	0.882 ± 0.001
ESMM + FR	0.899 ± 0.004	0.891 ± 0.002
ESMM + FR + LR	0.896 ± 0.006	0.888 ± 0.001

- ✓ FR (H1-only) < FR (H2-only) < LR
  - higher-level abstraction is more critical
- ✓ ESMM < ESMM + FR < LR
  - additive is better than multiplicative
- ✓ FR (H2-only) < FR (i.e., H1 + H2) < FR + LR
  - residual link in every layer benefits

Extension

Table 6: Performance on KuaiRand-Pure-S1 in terms of AUC. The bests are in bold font. The second bests are underlined.

Target	is_follow	is_forward	is_comment
NSE	0.821 ± 0.005	0.750 ± 0.011	0.779 ± 0.003
AITM	0.785 ± 0.014	<u>0.764 ± 0.012</u>	0.782 ± 0.005
ESMM	<u>0.822 ± 0.007</u>	0.757 ± 0.007	0.782 ± 0.004
$ESCM^2$-IPW	0.817 ± 0.010	0.754 ± 0.008	0.773 ± 0.004
$ESCM^2$-DR	0.821 ± 0.009	0.754 ± 0.008	0.782 ± 0.003
DCMT	0.819 ± 0.007	0.762 ± 0.008	<u>0.783 ± 0.003</u>
ResFlow	**0.826 ± 0.002**	**0.776 ± 0.006**	**0.789 ± 0.003**

- ResFlow generalizes well to non-sequentially dependent tasks.
- Building the task topology according to their sample sparsities is a good choice.

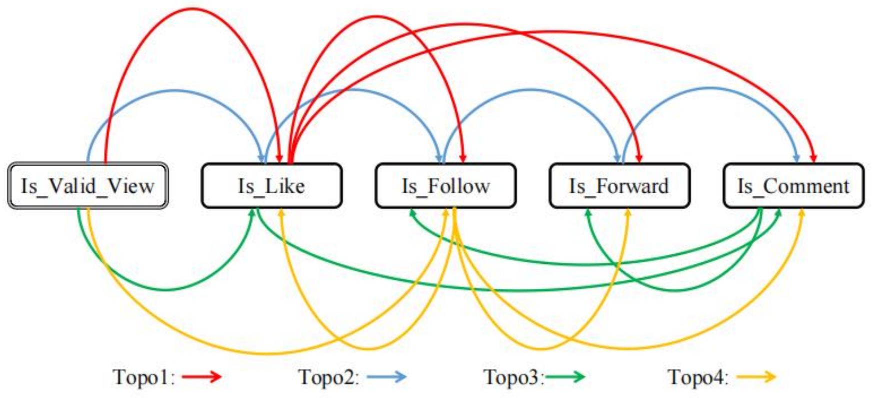


Figure 5: Visualization of different task topologies of KuaiRand-Pure-S1 Multi-Task.

Table 7: Performance on KuaiRand-Pure-S1 in terms of AUC. The bests are in bold font. The second bests are underlined.

Target	is_follow	is_forward	is_comment
ResFlow-topo1	**0.826 ± 0.002**	**0.776 ± 0.006**	**0.789 ± 0.003**
ResFlow-topo2	<u>0.825 ± 0.003</u>	0.756 ± 0.010	0.771 ± 0.002
ResFlow-topo3	0.811 ± 0.007	<u>0.774 ± 0.007</u>	0.778 ± 0.003
ResFlow-topo4	0.820 ± 0.002	0.748 ± 0.009	0.775 ± 0.002
NSE	0.821 ± 0.005	0.750 ± 0.011	<u>0.779 ± 0.003</u>

Extension

Regression (e.g., rating) as Progressive Multi-Task

- Task 0: the likelihood of v being larger than v_0
- Task 1: the likelihood of v being larger than v_1
- …
- Task K: the likelihood of v being larger than v_K

$$v_0 < v_2 < \cdots < v_K$$

Table 8: Regression performances on KuaiRand-Pure-S1 and MovieLens-1M in terms of MSE. Progressive indicates converting the regression into a progressive multi-task problem.

Dataset	KuaiRand-Pure-S1	MovieLens-1M
Traditional	1719.92 ± 3.82	0.906 ± 0.002
Progressive + NSE	1720.12 ± 4.32	0.906 ± 0.002
Progressive + ResFlow	**1658.44 ± 5.21**	**0.894 ± 0.002**

- ResFlow can also be used in regression tasks

04 Online Deployment

Score Fusion

How to fuse the predictions from multiple tasks into a simple ranking indicator:

- traditional, multiplicative: $\boldsymbol{CTR^{\alpha} * CTCVR^{\beta}}$
- proposed, additive: $\boldsymbol{\alpha * CTV + \beta * CTCVR}$

Method & Score Formula	WR@100	List AUC	OPU	CTR	CTCVR	BCR@20
ESMM, $CTR^{0.9} \times CVR^{1.1}$	+0.77%	+0.23%	+0.56%	-0.02%	+0.61%	+0.02%
ESMM, CTR*2+CTCVR*19	+2.7%	+0.09%	+0.75%	+0.12%	+0.67%	-0.01%
ResFlow, CTCVR	+1.41%	+1.82%	+0.45%	-0.15%	+0.40%	-0.02%
ResFlow, $CTR^{-0.2} \times CTCVR$	+3.37%	+1.03%	+1.33%	+0.06%	+1.19%	-0.02%
ResFlow, CTR+CTCVR*20	+4.19%	+0.88%	+2.14%	+0.66%	+2.11%	+0.02%
ResFlow, CTR+CTCVR*20+RL	+4.11%	+0.82%	+2.04%	+0.61%	+1.97%	-2.13%
ResFlow, CTR+CTCVR*20+RS	+0.03%	-0.13%	-0.53%	-0.19%	-0.44%	-2.99%

- Score fusion leads to better online performance than simply using CTCVR.
- Additive fusion leads to better online performance than multiplicative fusion.

Online-Offline Metric Alignment

Online metric OPU (order-per-user) is totally different from what we predict, while online A/B test is costly

- need a surrogate for online metric to more efficiently tune the model and fusion parameters
- it is a long-standing challenge that still remains unresolved

Propose an offline surrogate metric:

$$WR@K = \frac{\sum_{k=0}^{K} W_k}{\sum_{n=0}^{N} W_n},$$

- W_k indicates the number of orders of item k

Table 10: Pearson correlation coefficient (PCC) between offline metric uplift and online metric uplift.

Metrics	2 Targets Uplift PCC	2 Targets Uplift p-value	3 Targets Uplift PCC	3 Targets Uplift p-value
Recall@100	0.0814	0.4425	0.1974	0.0606
NDCG	0.3719	0.0002	0.3825	0.0002
List AUC	0.4353	1.6×10^{-5}	0.1564	0.1386
WR@100	0.7879	1.9×10^{-20}	0.8666	2.5×10^{-42}

- Significantly better than existing surrogates

Online Performance

Compared with the previously deployed:

- OPU: ↑1.29%
- Online CTR: ↑0.25%
- Online CTCVR: ↑1.37%

THANKS FOR YOUR ATTENTION!

Q & A

Based on joint work with Cong Fu et al. Deployed in Shopee Search.

Zhiming Zhou | SUFE | 2024/12/07