Recent RS papers

KDD, WWW, AAAI (2020)

Paper list (partial)

KDD:

- Multitask Mixture of Sequential Experts for User Activity Streams
- *Privileged Features Distillation at Taobao Recommendations*
- *Towards Automated Neural Interaction Discovery for Click-Through Rate Prediction*

WWW:

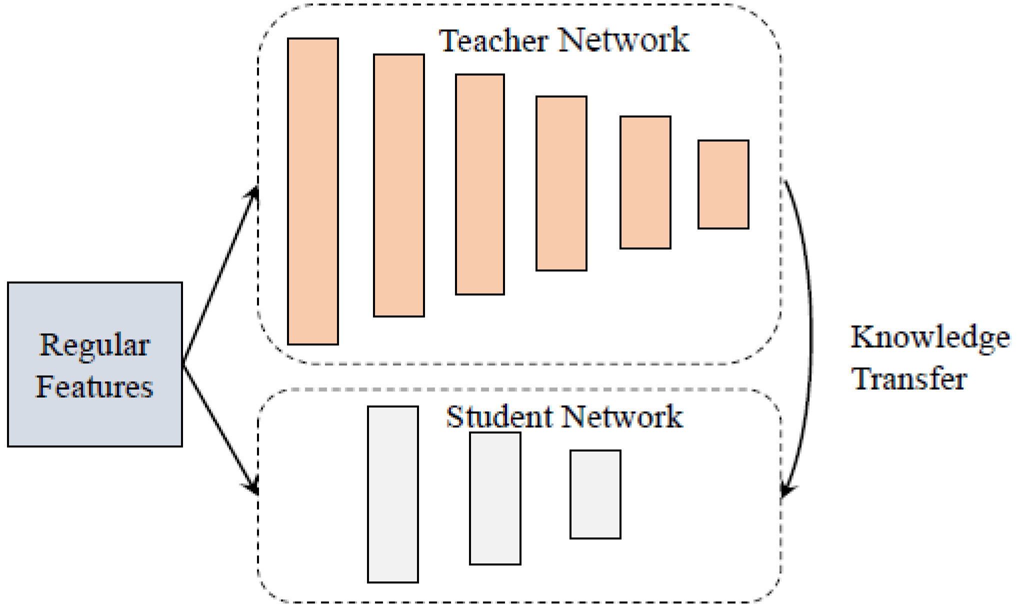
- Jointly Learning to Recommend and Advertise
- Adversarial Multimodal Representation Learning for Click-Through Rate Prediction
- *Beyond Clicks: Modeling Multi-Relational Item Graph for Session-Based Target Behavior Prediction (ECNU)*

AAAI:

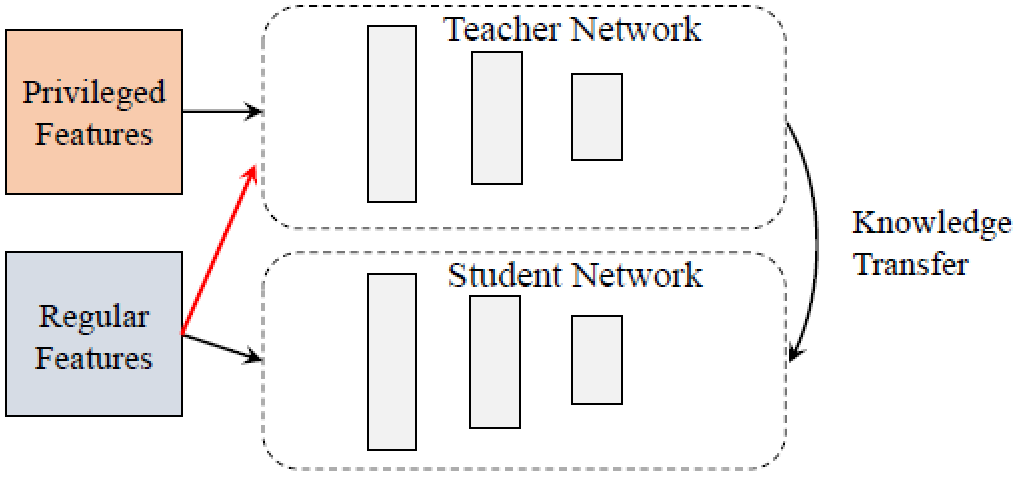
- Memory Augmented Graph Neural Networks for Sequential Recommendation
- Deep Time-Stream Framework for Click-Through Rate Prediction by Tracking Interest Evolution
- PEIA: Personality and Emotion Integrated Attentive Model for Music Recommendation on Social Media Platforms

Privileged Features Distillation at Taobao Recommendations

- Model Distillation & Privileged Features Distillation



(a) Model Distillation



(b) Privileged Features Distillation

Privileged Features Distillation at Taobao Recommendations

- Loss of Model Distillation

$$\min_{W_s} (1-\lambda) * L_s (\boldsymbol{y}, f_s(X; \boldsymbol{W}_s)) + \lambda * L_d (f_t(X; \boldsymbol{W}_t), f_s(X; \boldsymbol{W}_s))$$

- Loss of Learning using Privileged Information

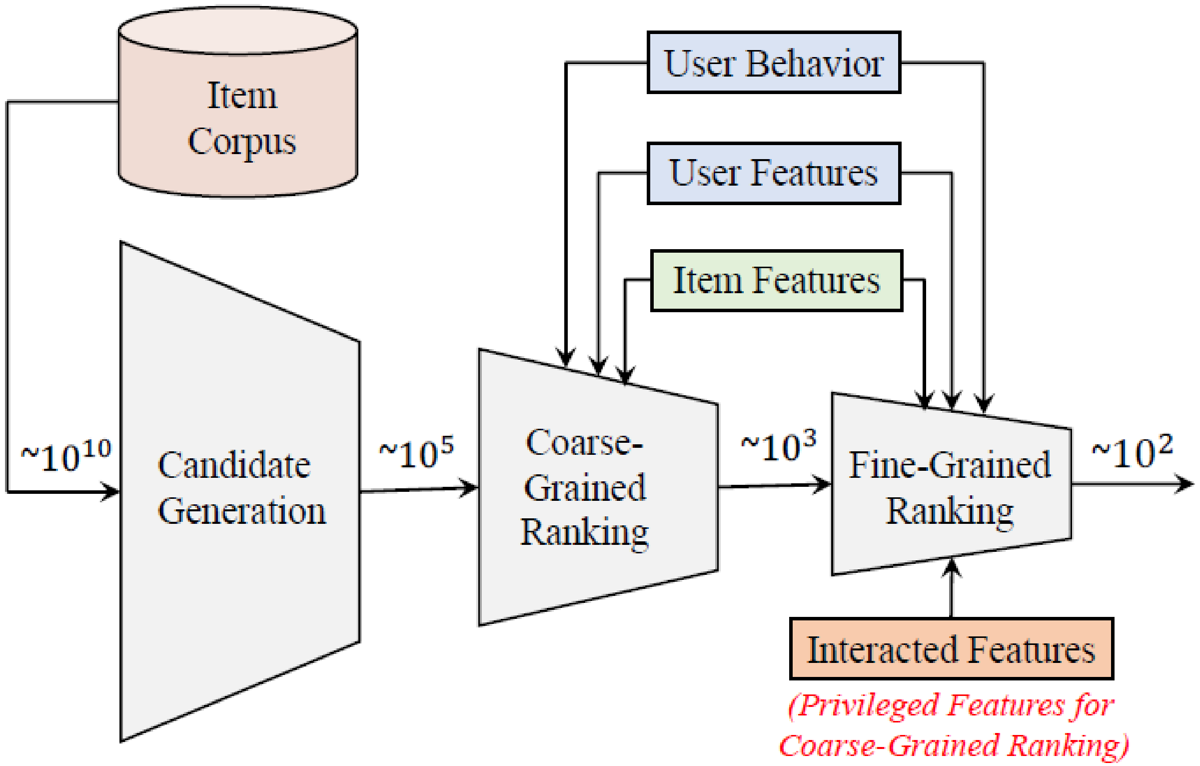
$$\min_{W_s} (1-\lambda) * L_s (\boldsymbol{y}, f(X; \boldsymbol{W}_s)) + \lambda * L_d \left(f(X^*; \boldsymbol{W}_t), f(X; \boldsymbol{W}_s)\right)$$

- Loss of Privileged Features Distillation

$$\min_{W_s} (1-\lambda) * L_s (\boldsymbol{y}, f(X; \boldsymbol{W}_s)) + \lambda * L_d \left(f(X, X^*; \boldsymbol{W}_t), f(X; \boldsymbol{W}_s)\right)$$

Privileged Features Distillation at Taobao Recommendations

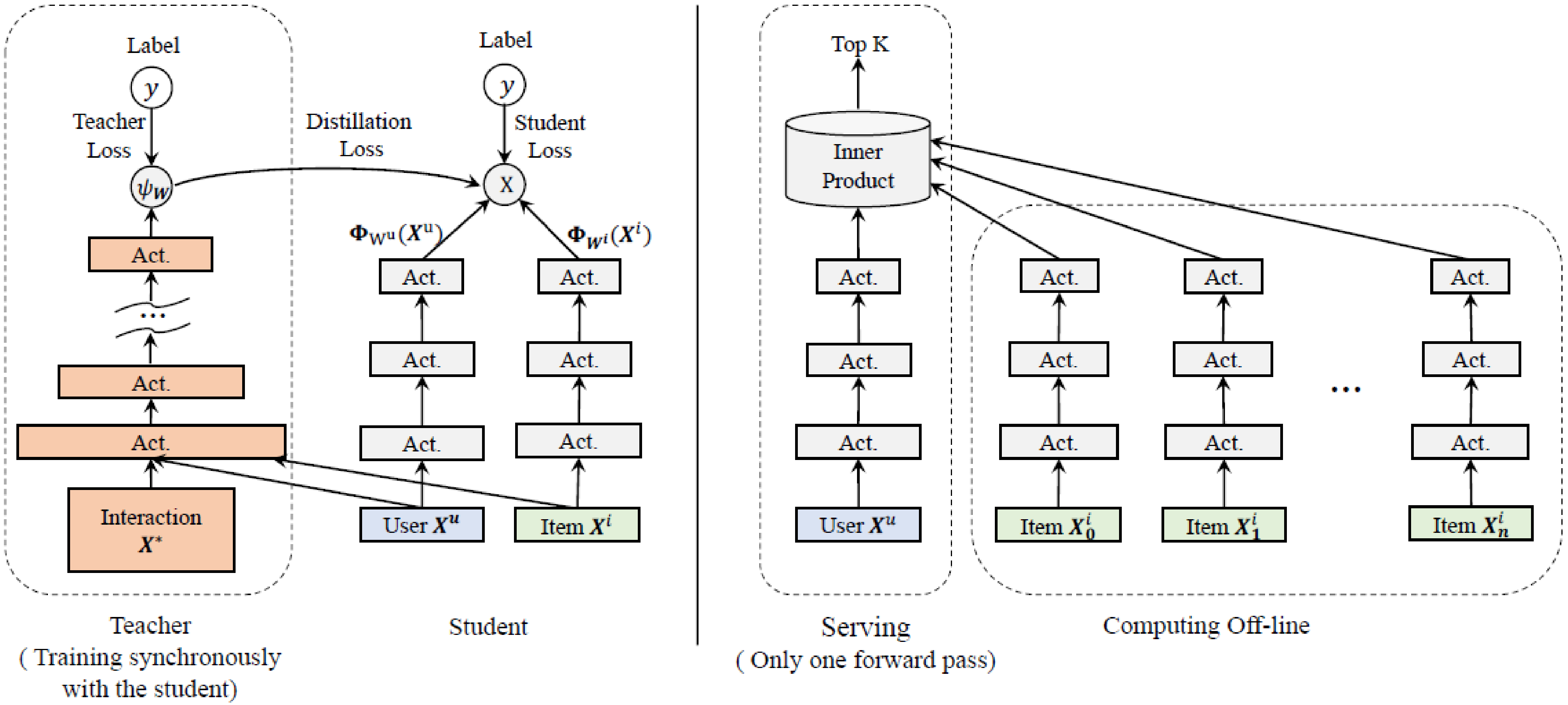
- Taobao Recommendations Procedure



Privileged Features Distillation at Taobao Recommendations

- An Illustration of Privileged Features

Privileged Features Distillation at Taobao Recommendations

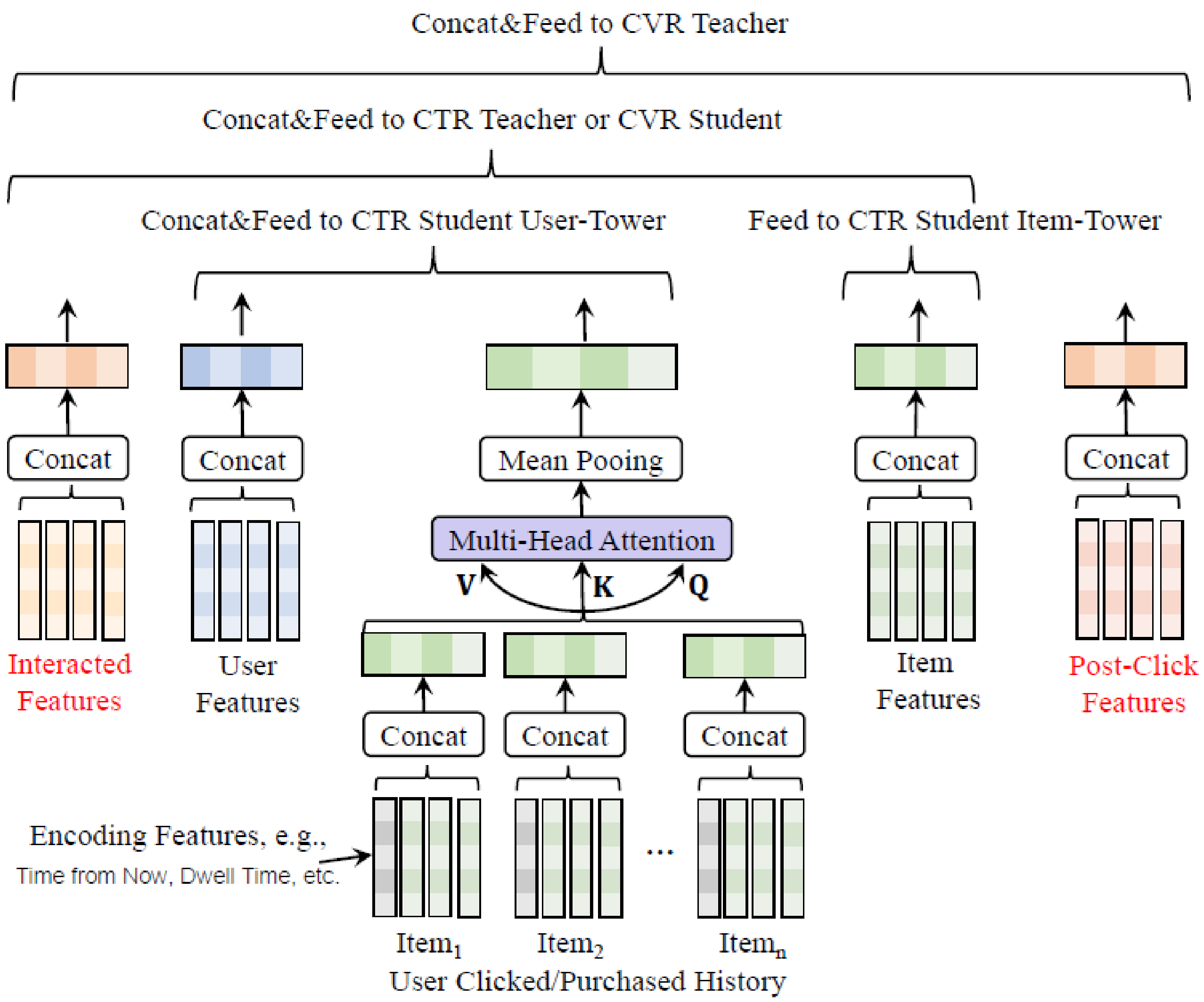


(a) Training with PFD+MD

(b) Serving with the Student

Privileged Features Distillation at Taobao Recommendations

- Experiment Setting



Privileged Features Distillation at Taobao Recommendations

- Results(AUC)

CTR

Methods	Dataset of 1 Day		Dataset of 10 Days	
	Student	Teacher	Student	Teacher
Baseline	0.6625	–	0.7042	–
LUPI [24]	0.6637	0.6687	–	–
MD [13]	0.6704	0.6892	–	–
PFD	0.6712	0.6921	–	–
PFD+MD	0.6745	0.7110	0.7160	0.7411

CVR

Methods	Dataset of 30days		Dataset of 60days	
	Student	Teacher	Student	Teacher
Baseline	0.9040	–	0.9082	–
MTL [31]	0.9045	–	0.9077	–
LUPI [24]	0.8965	0.9651	0.9003	0.9659
MD [13]	0.9052	0.9058	0.9093	0.9103
PFD	0.9084	0.9901	0.9135	0.9923
PFD+MD	0.9082	0.9911	0.9138	0.9929

Privileged Features Distillation at Taobao Recommendations

- Training Effects (Sync/Async Training, Independent/Shared Embedding)

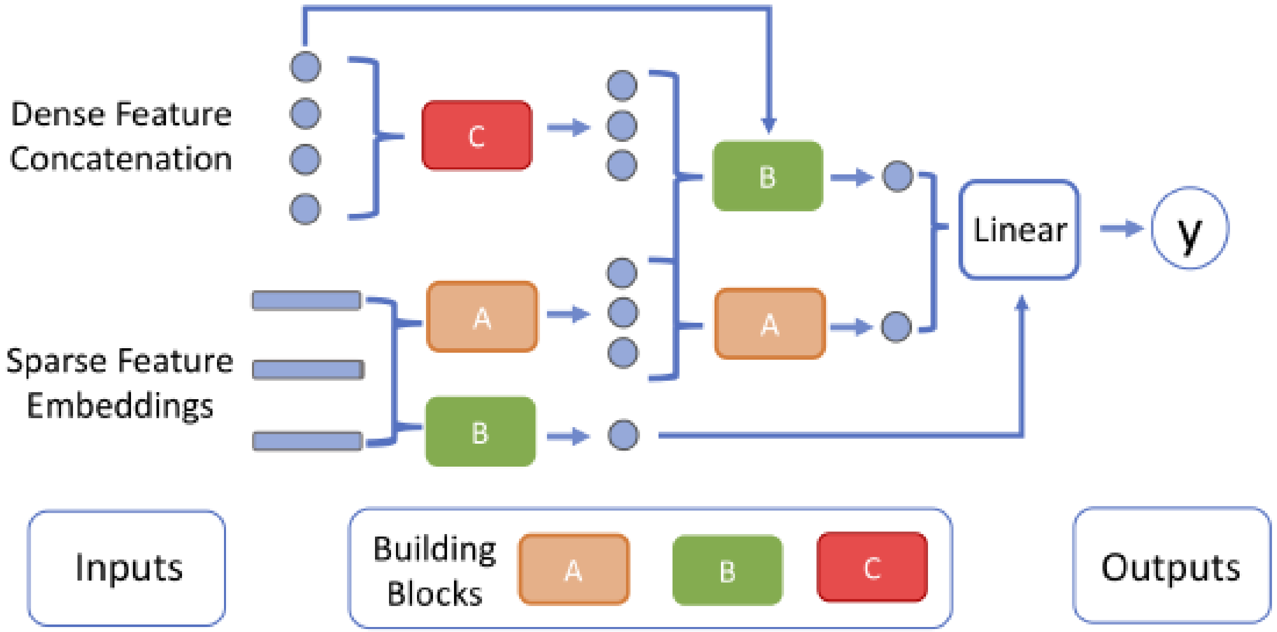
CTR

	Student	Teacher	Time	Relative
Baseline	0.6625	–	9.24 h	0%
Ind&Async	0.6751	0.7112	18.43 h	+99.5%
Ind&Sync	0.6748	0.7112	14.32 h	+55.0%
Share&Sync	0.6717	0.7108	9.51 h	+2.9%
Share*&Sync	0.6745	0.7110	10.29 h	+11.4%

CVR

	Student	Teacher	Time	Relative
Baseline	0.9040	–	12.22 h	0%
Ind&Async	0.9067	0.9887	26.85 h	+119.7%
Ind&Sync	0.9069	0.9887	20.56 h	+67.4%
Share&Sync	0.9082	0.9911	14.97 h	+22.5%
Share&Sync[†]	0.9084	0.9901	12.67 h	+3.6%

Towards Automated Neural Interaction Discovery for CTR Prediction



- NAS+CTR
- Block Type:
  MLP, FM, DP.
- Raw Feature Input Selection:
  Dense, Sparse, None, Both
- Inter-Block Connection
- Block Appendant Hyperparameters:
  The number of hidden units of MLP block

Towards Automated Neural Interaction Discovery for CTR Prediction

- Multi-Objective Survivor Selection

  Age, Performance, Complexity

$$f(q, a_A, r_A^q, c_A^q) = \mathbb{1}_{[a_A \leq q]} \cdot (\mu_1 a_A + \mu_2 r_A^q + \mu_3 c_A^q),$$

- Rank-Based Parent Selection

$$p(r_A^*) = \frac{\binom{r_A^* + \lambda - 1}{\lambda}}{\binom{p+\lambda}{1+\lambda}}, \qquad r_A^* \in \{1, 2, \ldots, p\}, \lambda \in \mathbb{N}^0,$$

- Guided Mutation by Learning to Hyperrank
  - Guider – GDBT
  - Random Neigbours
  - Learning to rank - Pairwise

Towards Automated Neural Interaction Discovery for CTR Prediction

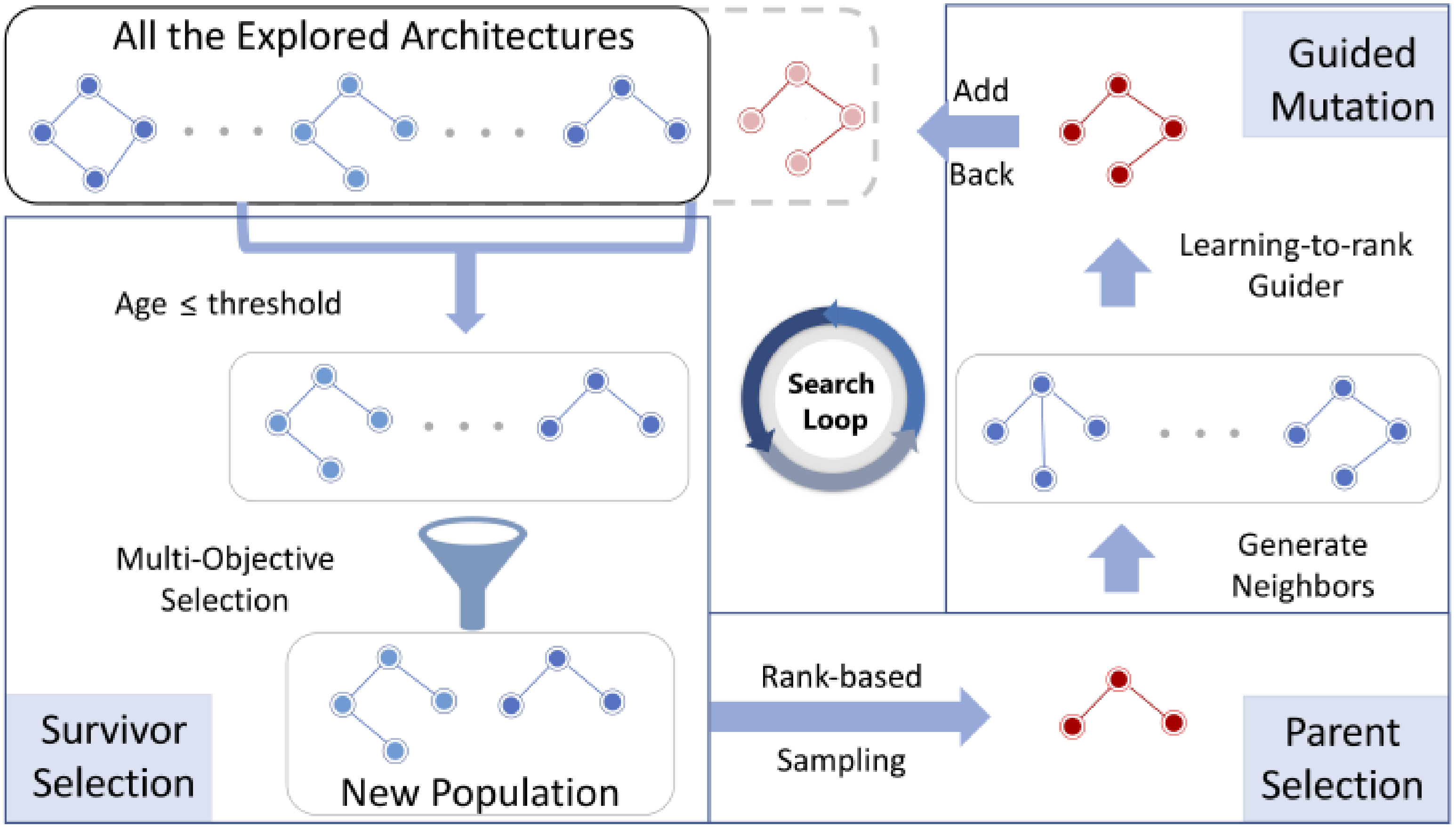


Figure 2: An illustration of the AutoCTR search loop

Towards Automated Neural Interaction Discovery for CTR Prediction

Table 1: General CTR prediction results on the three benchmark datasets

		Criteo		Avazu		KDD		Search cost
		Logloss	AUC	Logloss	AUC	Logloss	AUC	(GPU Days)
SOFA human-crafted Networks	DeepFM	0.4432	0.8086	0.3816	0.7767	0.1529	0.7974	-
	DLRM	0.4436	0.8085	0.3814	0.7766	0.1523	0.8004	-
	AutoInt+	0.4427	0.8090	0.3813	0.7772	0.1523	0.8002	-
Best Networks Found by the NAS Methods	Random	0.4421 ± 0.0003	0.8096 ± 0.0004	0.3824 ± 0.0030	0.7765 ± 0.0029	0.1531 ± 0.0001	0.8001 ± 0.0003	~ 0.75
	RL	0.4422 ± 0.0005	0.8094 ± 0.0005	0.3810 ± 0.0003	0.7778 ± 0.0005	0.1531 ± 0.0001	0.7999 ± 0.0002	~ 0.75
	LaNAS	0.4421 ± 0.0004	0.8096 ± 0.0005	0.3814 ± 0.0006	0.7772 ± 0.0011	0.1533 ± 0.0002	0.8001 ± 0.0009	~ 5
	LaNAS+	0.4417 ± 0.0001	0.8101 ± 0.0000	0.3800 ± 0.0004	0.7790 ± 0.0007	0.1521 ± 0.0001	0.8009 ± 0.0004	~ 0.75
	AutoCTR	**0.4413 ± 0.0002**	**0.8104 ± 0.0003**	**0.3800 ± 0.0001**	**0.7791 ± 0.0001**	**0.1520 ± 0.0000**	**0.8011 ± 0.0001**	~ 0.75
	AutoCTR (warm)	0.4417 ± 0.0005	0.8099 ± 0.0005	0.3804 ± 0.0004	0.7784 ± 0.0006	0.1523 ± 0.0001	0.8004 ± 0.0003	~ 0.75

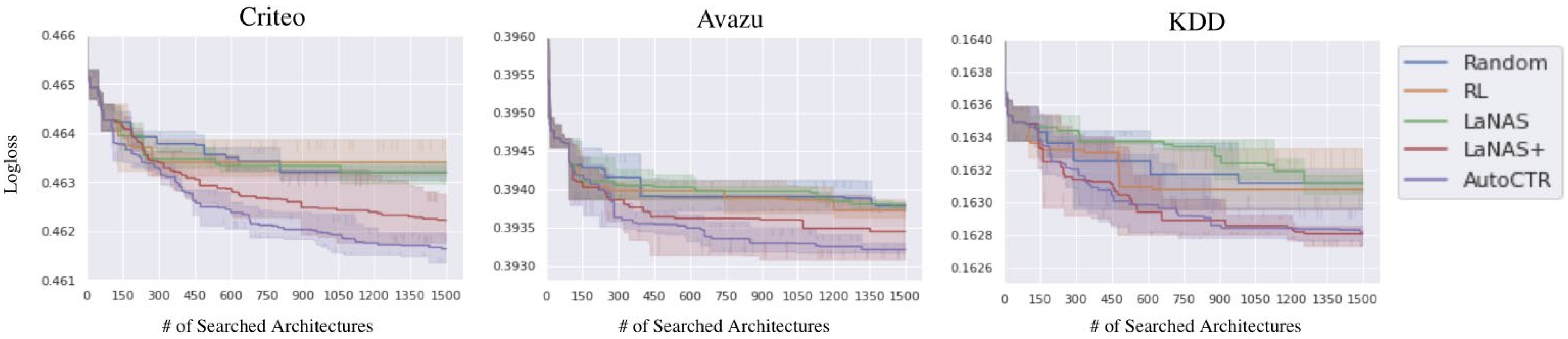


Figure 4: The performance drifting of the best architecture during the search process on different datasets

Towards Automated Neural Interaction Discovery for CTR Prediction

Table 3: Transferability of architectures found by AutoCTR

Original Dataset \ Target Dataset	Criteo	Avazu	KDD
Criteo	0.4413	0.3799	0.1520
Avazu	0.4421	0.3800	0.1535
KDD	0.4418	0.3803	0.1521

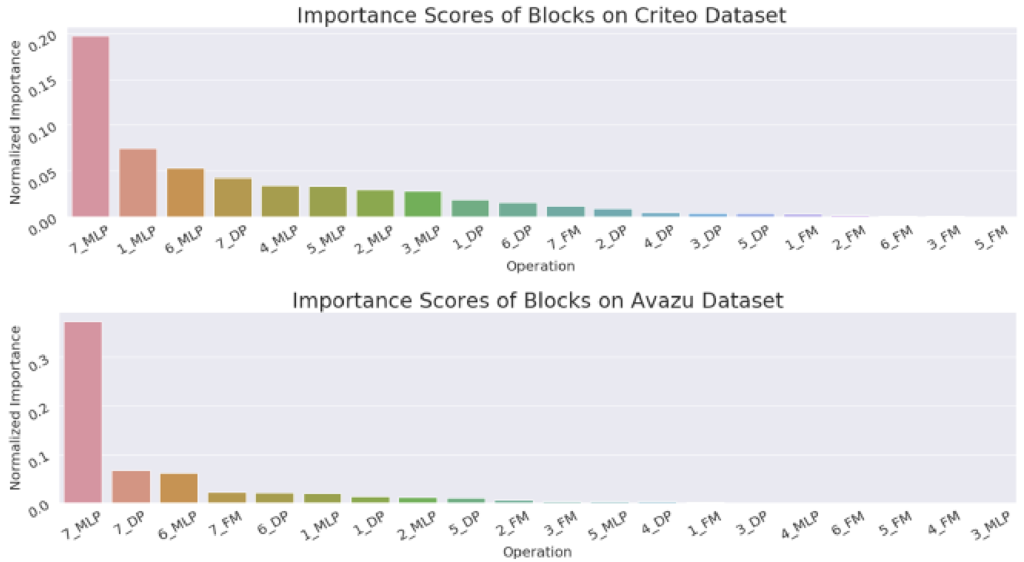


Figure 7: Normalized importance scores of top-20 block type operations learned by AutoCTR guider on Criteo and Avazu

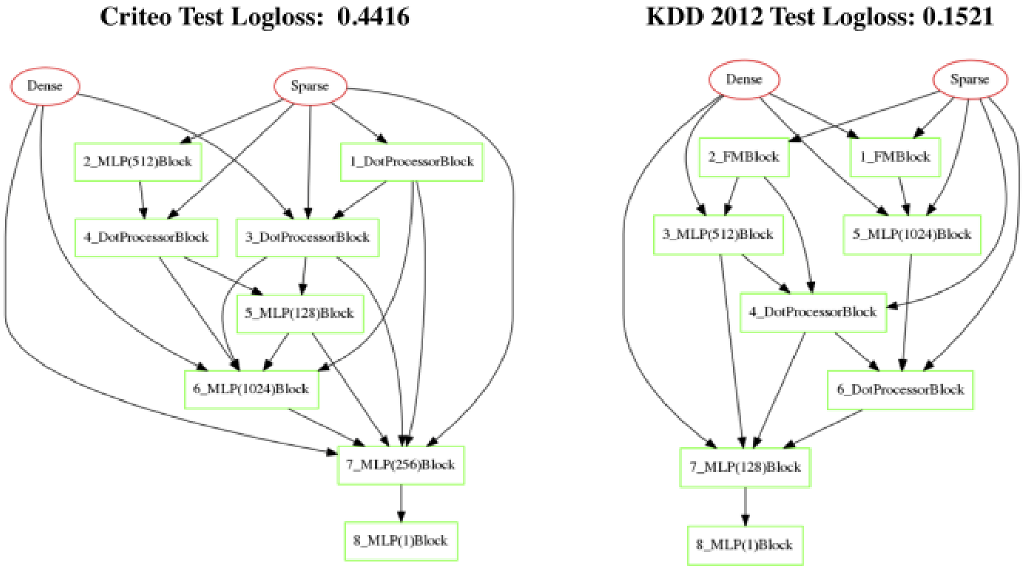


Figure 6: Two architectures found by AutoCTR.

Beyond Clicks: Modeling Multi-Relational Item Graph for Session-Based Target Behavior Prediction

- Limitations:

1. “only utilizing the same type of user behavior for prediction, but ignore the potential of taking other behavior data as auxiliary information”
2. “item-to-item relations are modeled separately and locally in one behavior sequence”

- GNN

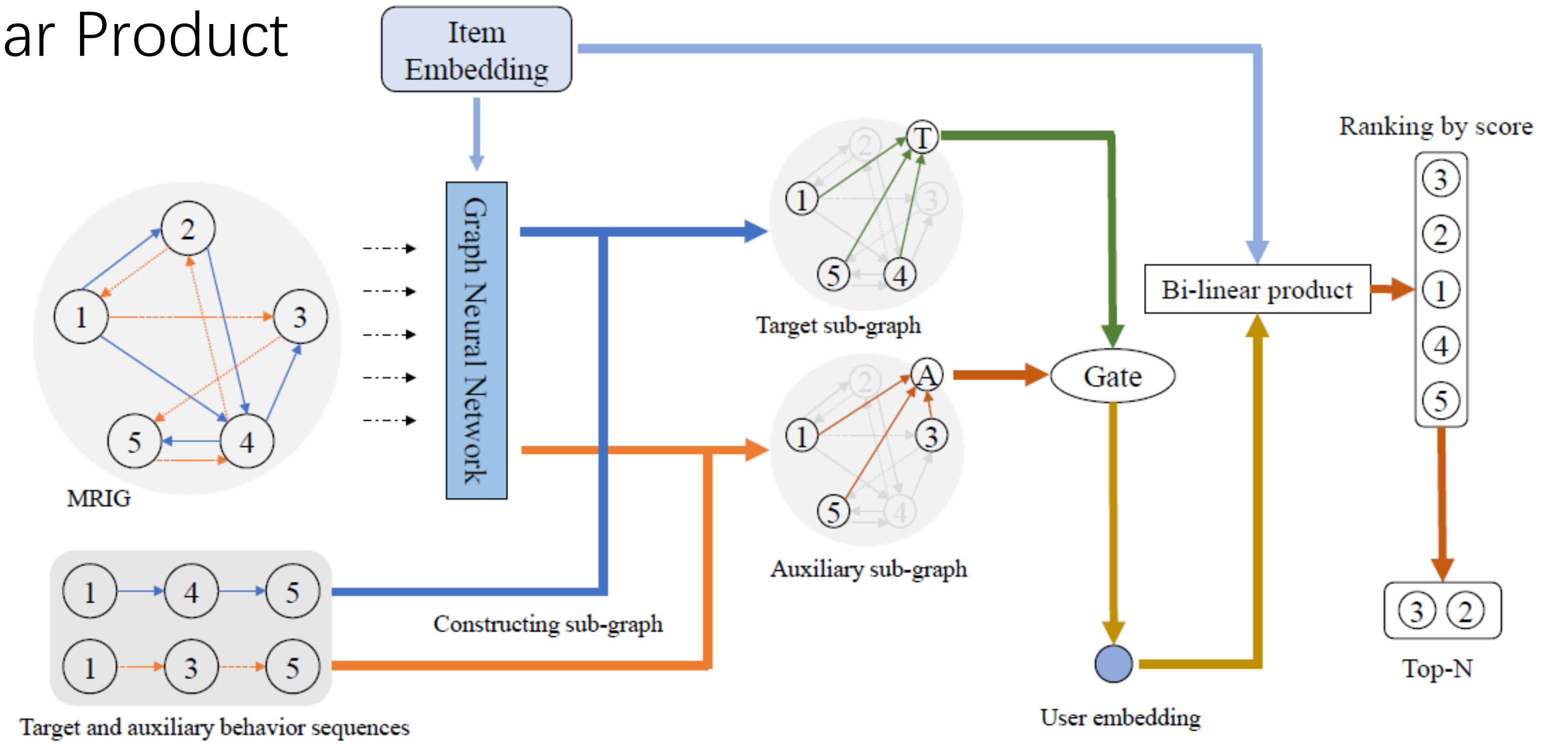
Beyond Clicks: Modeling Multi-Relational Item Graph for Session-Based Target Behavior Prediction

- Architecture
  - MRIG - Item Representation Learning
  - Virtual Node- Sequence Representation Learning
  - Scoring – Gate & Bilinear Product

$$\alpha = \sigma(\mathbf{W}_g[\mathbf{p}; \mathbf{q}]),$$

$$\mathbf{o} = \alpha \cdot \mathbf{p} + (1 - \alpha) \cdot \mathbf{q}.$$

$$s_v = \mathbf{o}^\top \mathbf{W} \mathbf{e}_v,$$



Beyond Clicks: Modeling Multi-Relational Item Graph for Session-Based Target Behavior Prediction

- Data

Table 1: Basic statistics of the datasets.

Data	WeChat	Yoochoose
#items	56,561	52,740
#sessions	100,000	9,249,729
Time duration	2019/09/17~23	2014/04/01~09/30
#edge of target	217,774	225,879
#edge of auxiliary	1,546,220	3,277,411
Average length of target	9.76	3.31
Average length of auxiliary	33.49	8.56
#training data	167,931	163,005
#validation data	12,333	12,985
#test data	24,667	25,971

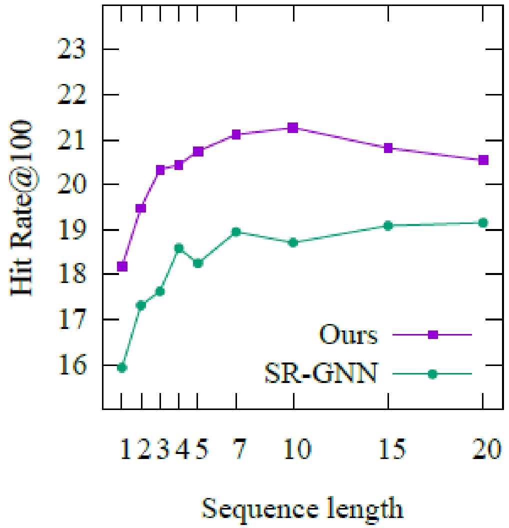
Beyond Clicks: Modeling Multi-Relational Item Graph for Session-Based Target Behavior Prediction

- Results

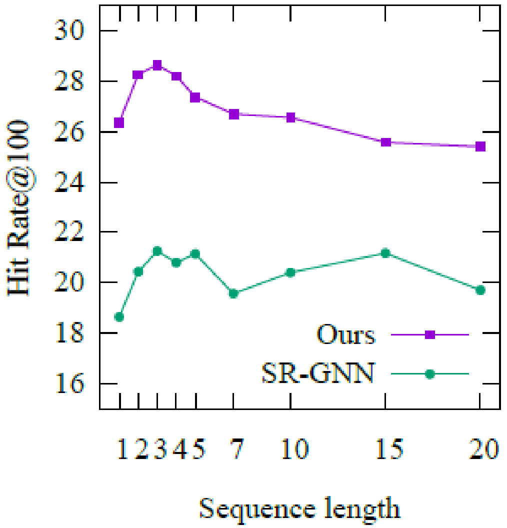
Methods	WeChat			Yoochoose		
	H@100	M@100	N@100	H@100	M@100	N@100
POP	13.565	1.1247	3.2621	6.095	0.2529	1.2231
Item-KNN	15.770	1.1624	3.7222	15.286	1.9415	4.4040
GRU4Rec	18.831	1.3956	4.3966	19.114	2.5292	5.5830
NARM	19.131	1.4034	4.4416	18.775	2.5819	5.5813
STAMP	17.757	1.3083	4.1078	20.361	2.3487	5.6879
SR-GNN	18.940	1.3827	4.3967	21.262	2.6892	6.1232
GC-SAN	19.034	1.2090	4.2490	19.718	2.5218	5.6861
HetGNN	20.290	1.4171	4.6504	24.031	2.9546	6.8732
R-DAN	18.952	1.3879	4.3980	15.956	2.3107	4.8608
CoAtt	17.700	1.1931	4.0137	20.080	2.5742	5.8206
Ours	**21.271**	**1.4797**	**4.8529**	**28.632**	**3.6564**	**8.2722**
GRU4Rec (w/o a)	16.889	1.2346	3.9128	14.817	1.6032	4.0012
NARM (w/o a)	17.773	1.3123	4.1298	14.443	1.5540	3.8900
SR-GNN (w/o a)	18.093	1.2621	4.1368	15.302	1.5782	4.0852
Ours (w/o a)	19.252	1.3933	4.4473	21.089	2.3798	5.8221

Table 4: Ablation study of MGNN-SPredl.

Methods	WeChat			Yoochoose		
	H@100	M@100	N@100	H@100	M@100	N@100
Ours (w/o ae)	20.923	1.4665	4.7945	25.463	2.7678	6.8907
Ours (w/o asg)	19.742	1.3949	4.5167	22.517	2.6025	6.2631
Ours (w/o g)	20.363	1.3707	4.6154	27.577	3.3531	7.7896
Ours	**21.271**	**1.4797**	**4.8529**	**28.632**	**3.6564**	**8.2722**



(a) WeChat



(b) Yoochoose