CL4CTR: A Contrastive Learning Framework for CTR Prediction

Fangye Wang*
School of Computer Science
Fudan University, Shanghai, China
fywang18@fudan.edu.cn

Yingxu Wang*
School of Computer Science
Fudan University, Shanghai, China
yingxuwang20@fudan.edu.cn

Dongsheng Li
Microsoft Research Asia
Shanghai, China
dongsli@microsoft.com

Hansu Gu†
Seattle, United States
hansug@acm.org

Tun Lu*†
School of Computer Science
Fudan University, Shanghai, China
lutun@fudan.edu.cn

Peng Zhang*
School of Computer Science
Fudan University, Shanghai, China
zhangpeng_@fudan.edu.cn

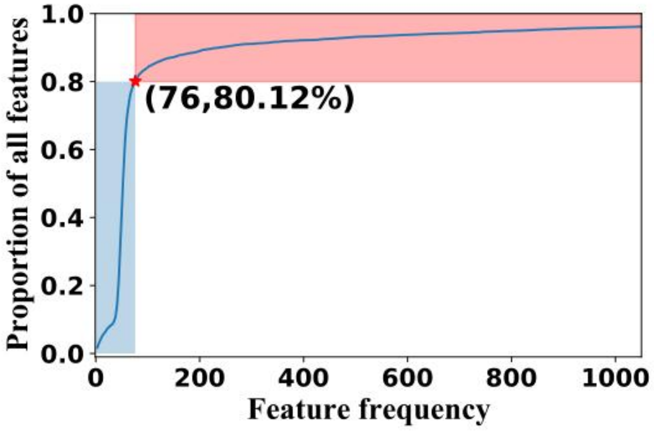
Ning Gu*
School of Computer Science
Fudan University, Shanghai, China
ninggu@fudan.edu.cn

code: https://github.com/cl4ctr/cl4ctr

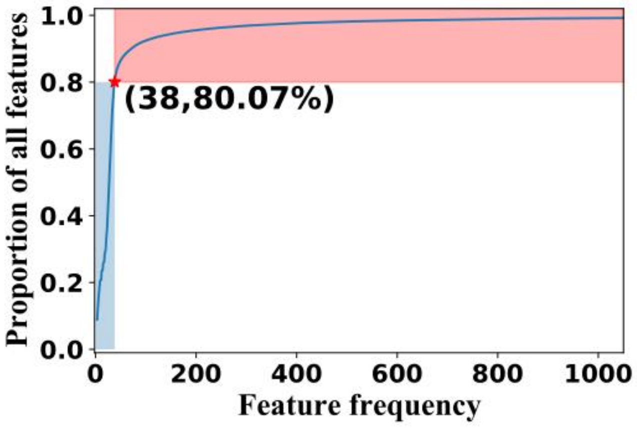
WSDM 2023

Reported by Minqin Li

Introduction



(a) Frappe



(b) ML-tag

Figure 1: Cumulative distribution of feature frequencies. (38, 80.07%) indicates that features with feature frequencies less than or equal to 38 times account for 80.07% of all features.

Many Click-Through Rate prediction works focused on designing advanced architectures to model complex feature interactions but neglected the importance of feature representation learning.

High frequency features have higher chances to be trained than low frequency features, causing the representations of low frequency features to be sub-optimal.

Method

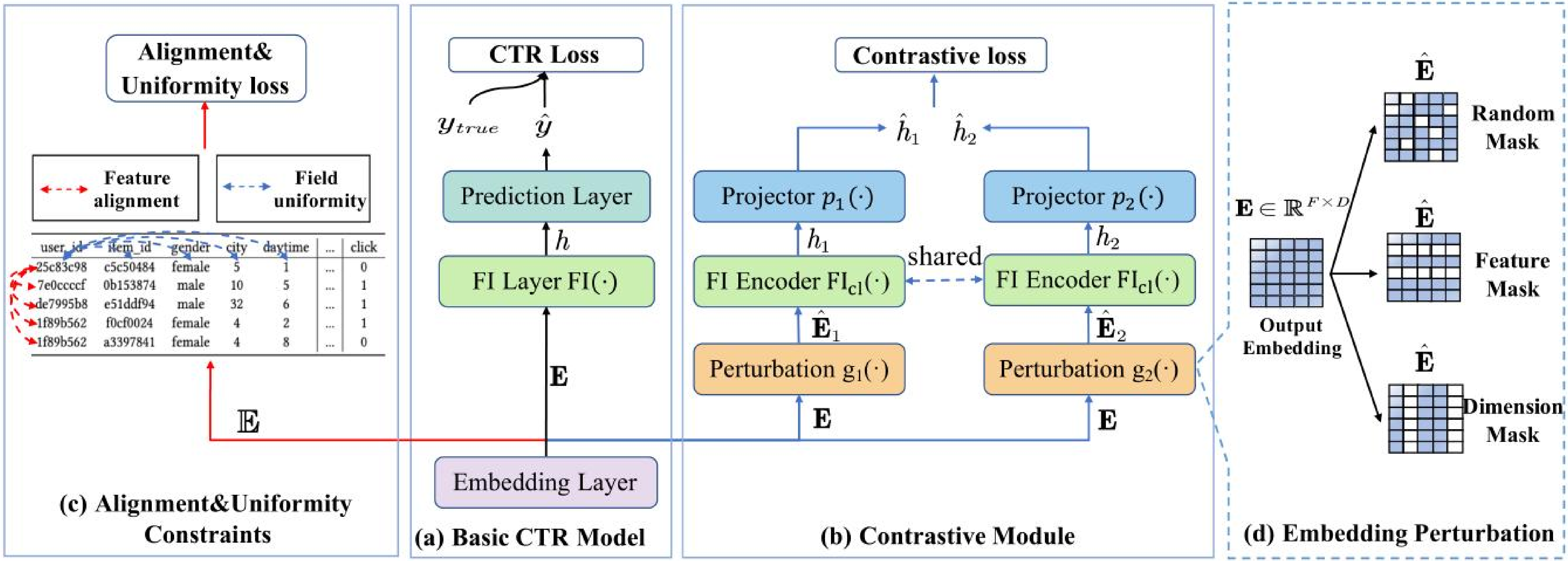
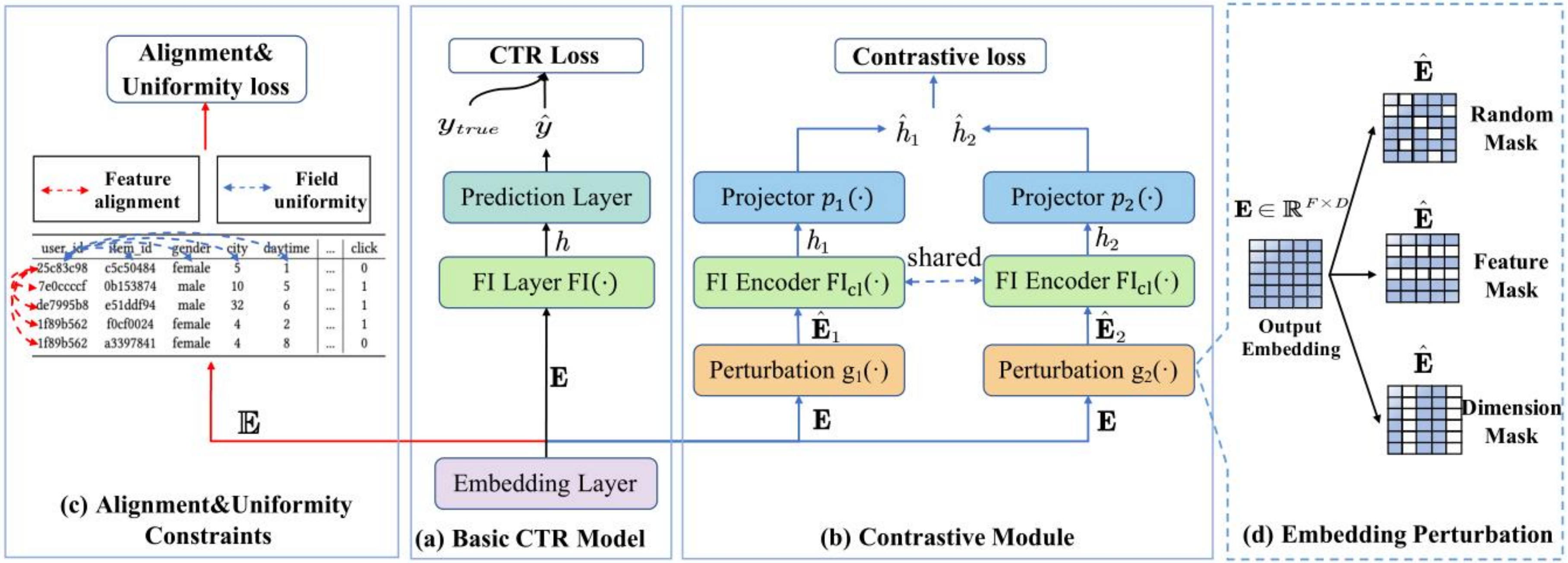


Figure 2: Architecture of the CL4CTR framework. CL4CTR including three components: (a) a basic CTR model; (b) a contrastive module; (c) alignment & uniformity constraints. In contrastive module, we design (d) three embedding perturbation methods.

Method



$$\hat{\mathrm{E}} = \mathrm{g}_r(\mathrm{E}) = \mathrm{E} \cdot \mathrm{I}, \mathrm{I} \sim \mathrm{Bernoulli}(p) \in \mathbb{E}^{F\times D} \tag{1}$$

$$\hat{\mathrm{E}} = \mathrm{g}_f(\mathrm{E}) = [\hat{\mathrm{e}}^1; \hat{\mathrm{e}}^2; ...; \hat{\mathrm{e}}^F], \hat{\mathrm{e}}^f = \begin{cases} \mathrm{e}^f, & t \notin \mathcal{T} \\ [\mathrm{mask}], & t \in \mathcal{T} \end{cases} \tag{2}$$

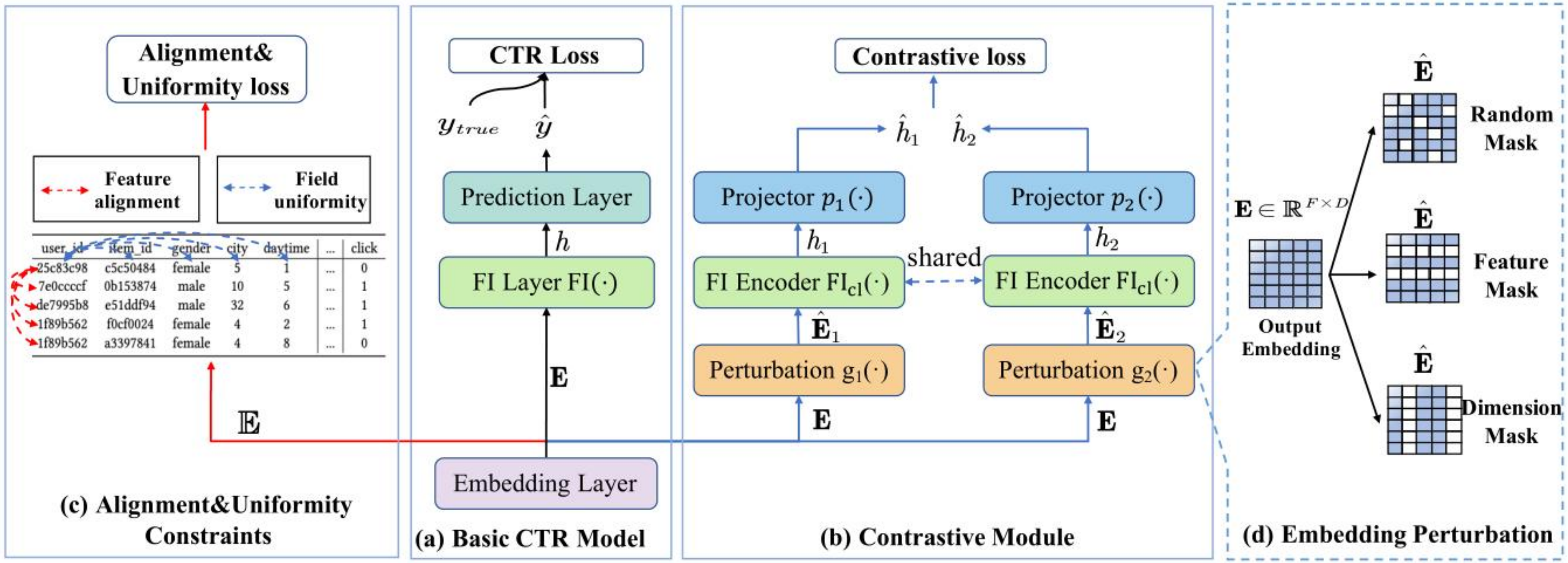
$$\hat{\mathrm{E}} = \mathrm{g}_d(\mathrm{E}) = [d\mathrm{e}^1; d\mathrm{e}^2; ...; d\mathrm{e}^F], d \sim \mathrm{Bernoulli}(p) \in \mathbb{R}^D \tag{3}$$

$$h_1 = FI_{cl}(\hat{\mathrm{E}}_1), h_2 = FI_{cl}(\hat{\mathrm{E}}_2) \tag{4}$$

$$\hat{h}_1 = p_1(h_1), \hat{h}_2 = p_2(h_2) \tag{5}$$

$$\mathcal{L}_{cl} = \frac{1}{B} \sum_{i=1}^{B} \left\| \hat{h}_{i,1} - \hat{h}_{i,2} \right\|_2^2 \tag{6}$$

Method



$$\mathcal{L}_{ctr} = -\frac{1}{N}\sum_{i=1}^{N}\left(y_i \log\left(\sigma\left(\hat{y}_i\right)\right) + \left(1-y_i\right)\log\left(1-\sigma\left(\hat{y}_i\right)\right)\right) \tag{7}$$

$$\mathcal{L}_a = \sum_{f=1}^{F}\sum_{\mathrm{e_i},\mathrm{e_j}\in \mathrm{E}_f} \|\mathrm{e_i}-\mathrm{e_j}\|_2^2 \tag{8}$$

$$\mathcal{L}_u = \sum_{\substack{\mathrm{e_i}\in \mathrm{E}_f \\ 1<=f<=F}} \sum_{\mathrm{e_j}\in(\mathbb{E}-\mathrm{E}_f)} sim(\mathrm{e_i},\mathrm{e_j}) \tag{9}$$

$$\mathcal{L}_{total} = \mathcal{L}_{ctr} + \alpha\cdot\mathcal{L}_{cl} + \beta\cdot\left(\mathcal{L}_a + \mathcal{L}_u\right) \tag{10}$$

Experiments

Table 1: An example of multi-field tabular data for CTR prediction. Each row represents an input instance and each column indicates a field. Moreover, each field contains multiple features, but each feature only belongs to one field.

user_id	item_id	gender	city	daytime	...	click
25c83c98	c5c50484	female	5	1	...	0
7e0ccccf	0b153874	male	10	5	...	1
de7995b8	e51ddf94	male	32	6	...	1
1f89b562	f0cf0024	female	4	2	...	1
1f89b562	a3397841	female	4	8	...	0

Experiments

Table 2: Dataset statistics.

Datasets	Positive	#Training	#Validation	#Test	#Features	#Fields
Frappe	33%	202K	58K	29K	5K	10
ML-tag	33%	1,404K	401K	201K	90K	3
ML-1M	57.5%	800K	100K	100K	10K	5
SafeDriver	3.64%	476K	59K	59K	600	57

Experiments

Model Class	Datasets / Model	Frappe AUC	Frappe Logloss	ML-tag AUC	ML-tag Logloss	ML-1M AUC	ML-1M Logloss	SafeDriver AUC	SafeDriver Logloss	ΔAUC ↑	ΔLogloss ↓
First-order	LR	0.9331	0.2894	0.9348	0.2960	0.7899	0.5417	0.6244	0.1622	-3.35%	0.0572
Second-Order	FM	0.9746	0.1856	0.9488	0.2595	0.8023	0.5332	0.6301	0.1538	-1.22%	0.0179
	FwFM	0.9756	0.1784	0.9582	0.2531	0.8046	0.5281	0.6335	0.1532	-0.74%	0.0131
	IFM	0.9771	0.1581	0.9515	0.2497	0.8080	0.5286	0.6353	0.1526	-0.70%	0.0071
	FmFM	0.9801	0.1682	0.9552	0.2493	0.8093	0.5264	0.6378	0.1518	-0.39%	0.0088
High-Order	CrossNet	0.9800	0.1658	0.9549	0.2480	0.8114	0.5218	0.6336	0.1517	-0.50%	0.0067
	IPNN	0.9809	0.1604	0.9607	0.2295	0.8110	0.5190	0.6373	0.1521	-0.19%	0.0001
	OPNN	0.9799	0.1683	0.9599	0.2421	0.8112	0.5185	0.6375	0.1519	-0.22%	0.0051
	FINT	0.9807	0.1578	0.9557	0.2430	0.8123	0.5192	0.6349	0.1522	-0.38%	0.0029
	DCAP	0.9801	0.1612	0.9560	0.2428	0.8130	0.5171	0.6390	0.1512	-0.20%	0.0030
Ensemble	WDL	0.9770	0.1783	0.9599	0.2660	0.8093	0.5226	0.6353	0.1525	-0.44%	0.0110
	DCN	0.9788	0.1621	0.9550	0.2472	0.8125	0.5175	0.6379	0.1514	-0.32%	0.0044
	DeepFM	0.9780	0.1732	0.9586	0.2551	0.8061	0.5259	0.6318	0.1529	-0.69%	0.0117
	xDeepFM	0.9799	0.1750	0.9604	0.2472	0.8082	0.5244	0.6403	0.1515	-0.19%	0.0094
	FiBiNET	0.9793	0.1707	0.9548	0.2532	0.8032	0.5313	0.6391	0.1505	-0.56%	0.0113
	AutoInt+	0.9783	0.1762	0.9535	0.2562	0.8099	0.5219	0.6310	0.1516	-0.73%	0.0114
	AFN+	0.9786	0.1637	0.9561	0.2468	0.8041	0.5304	0.6374	0.1517	-0.58%	0.0080
	TFNet	0.9798	0.1708	0.9527	0.2551	0.8099	0.5212	0.6387	0.1533	-0.41%	0.0100
	FED	0.9791	0.1606	0.9557	0.2465	0.8128	0.5184	0.6369	0.1534	-0.33%	0.0046
	DCN-V2	0.9803	0.1595	0.9610	0.2330	0.8132	0.5169	0.6406	0.1510	-	-
Ours	*CL4CTR$_{FM}$*	**0.9822**	**0.1324**	**0.9621**	**0.2102**	**0.8164**	**0.5136**	**0.6449**	**0.1483**	0.34%	-0.0140
	RelaImp	0.13%	16.10%	0.11%	8.41%	0.39%	0.64%	0.67%	1.46%	-	-

Table 3: Overall accuracy comparison in the four datasets.

Experiments

Table 4: Compatibility study of CL4CTR.

Model	Frappe		ML-1M		SafeDriver	
	AUC	Logloss	AUC	Logloss	AUC	Logloss
FM	0.9746	0.1856	0.8023	0.5332	0.6244	0.1622
$CL4CTR_{FM}$	**0.9822**	**0.1324**	**0.8164**	**0.5136**	**0.6449**	**0.1483**
FwFM	0.9756	0.1784	0.8046	0.5281	0.6335	0.1532
$CL4CTR_{FwFM}$	**0.9815**	**0.1532**	**0.8118**	**0.5192**	**0.6421**	**0.1487**
DeepFM	0.9780	0.1732	0.8061	0.5259	0.6318	0.1529
$CL4CTR_{DeepFM}$	**0.9813**	**0.1677**	**0.8113**	**0.5194**	**0.6381**	**0.1504**
Autoint+	0.9783	0.1762	0.8099	0.5219	0.6310	0.1516
$CL4CTR_{Autoint+}$	**0.9802**	**0.1684**	**0.8122**	**0.5174**	**0.6402**	**0.1506**
DCN	0.9788	0.1621	0.8125	0.5170	0.6379	0.1514
$CL4CTR_{DCN}$	**0.9808**	**0.1566**	**0.8164**	**0.5125**	**0.6415**	**0.1494**
DCN-V2	0.9803	0.1595	0.8132	0.5169	0.6406	0.1510
$CL4CTR_{DCN-V2}$	**0.9812**	**0.1549**	**0.8144**	**0.5153**	**0.6411**	**0.1497**

Experiments

Table 5: Impact of data augmentation methods.

Base model	Variants	Frappe AUC	Frappe Logloss	SafeDriver AUC	SafeDriver Logloss
FM	Base	0.9746	0.1856	0.6244	0.1622
	Random	**0.9822**	**0.1324**	**0.6449**	**0.1483**
	Feature	0.9814	0.1328	0.6303	0.1539
	Dimension	0.9816	0.1334	0.6404	0.1505
FwFM	Base	0.9756	0.1784	0.6335	0.1532
	Random	0.9815	0.1532	0.6421	0.1487
	Feature	**0.9822**	**0.1513**	0.6384	0.1483
	Dimension	0.9811	0.1465	**0.6455**	**0.1508**
DeepFM	Base	0.9780	0.1817	0.6318	0.1529
	Random	**0.9813**	**0.1677**	**0.6381**	**0.1504**
	Feature	0.9798	0.1750	0.6341	0.1522
	Dimension	0.9804	0.1697	0.6353	0.1514
DCN	Base	0.9788	0.1611	0.6379	0.1514
	Random	**0.9808**	**0.1566**	**0.6415**	**0.1494**
	Feature	0.9804	0.1601	0.6409	0.1508
	Dimension	0.9803	0.1573	0.6411	0.1504

Table 6: Impact of different FI encoder $FI_{cl}(\cdot)$.

Base model	FI Encoder	Frappe AUC	Frappe Logloss	ML-1M AUC	ML-1M Logloss
FM	Base	0.9746	0.1856	0.8023	0.5332
	DNN	0.9804	0.1404	**0.8177**	**0.5123**
	Transformer	**0.9822**	**0.1324**	0.8164	0.5136
	CrossNet2	0.9801	0.1438	0.8170	0.5143
FwFM	Base	0.9756	0.1784	0.8046	0.5281
	DNN	0.9809	**0.1504**	0.8064	0.5264
	Transformer	0.9815	0.1532	**0.8118**	**0.5192**
	CrossNet2	**0.9822**	0.1675	0.8102	0.5231
DeepFM	Base	0.9780	0.1732	0.8061	0.5259
	DNN	0.9804	0.1710	0.8101	0.5206
	Transformer	**0.9813**	**0.1704**	**0.8113**	**0.5194**
	CrossNet2	0.9791	0.1719	0.8109	0.5202
DCN-V2	Base	0.9803	0.1595	0.8132	0.5169
	DNN	0.9807	0.1573	**0.8151**	**0.5144**
	Transformer	**0.9812**	**0.1549**	0.8144	0.5153
	CrossNet2	0.9804	0.1588	0.8141	0.5155

Experiments

Table 7: Impact of SSL signals in the loss function.

Model	Loss Function	Frappe		ML-1M	
		AUC	Logloss	AUC	Logloss
FM	$\mathcal{L}_{ctr}$	0.9746	0.1856	0.8023	0.5332
	$+ \mathcal{L}_{cl}$	0.9794	0.1485	0.8102	0.5230
	$+ (\mathcal{L}_a + \mathcal{L}_u)$	0.9812	0.1455	0.8139	0.5175
	$+ \mathcal{L}_{cl} + (\mathcal{L}_a + \mathcal{L}_u)$	**0.9822**	**0.1324**	**0.8164**	**0.5136**
FwFM	$\mathcal{L}_{ctr}$	0.9756	0.1784	0.8046	0.5281
	$+ \mathcal{L}_{cl}$	0.9785	0.1553	0.8109	0.5229
	$+ (\mathcal{L}_a + \mathcal{L}_u)$	0.9812	0.1536	0.8098	0.5252
	$+ \mathcal{L}_{cl} + (\mathcal{L}_a + \mathcal{L}_u)$	**0.9815**	**0.1532**	**0.8118**	**0.5192**
DeepFM	$\mathcal{L}_{ctr}$	0.9780	0.1817	0.8061	0.5259
	$+ \mathcal{L}_{cl}$	0.9794	0.1701	0.8094	0.5235
	$+ (\mathcal{L}_a + \mathcal{L}_u)$	0.9784	0.1791	0.8103	0.5214
	$+ \mathcal{L}_{cl} + (\mathcal{L}_a + \mathcal{L}_u)$	**0.9813**	**0.1677**	**0.8113**	**0.5194**
DCN	$\mathcal{L}_{ctr}$	0.9788	0.1611	0.8125	0.5170
	$+ \mathcal{L}_{cl}$	0.9802	0.1585	0.8138	0.5150
	$+ (\mathcal{L}_a + \mathcal{L}_u)$	0.9792	0.1600	0.8129	0.5188
	$+ \mathcal{L}_{cl} + (\mathcal{L}_a + \mathcal{L}_u)$	**0.9808**	**0.1566**	**0.8164**	**0.5125**

Experiments

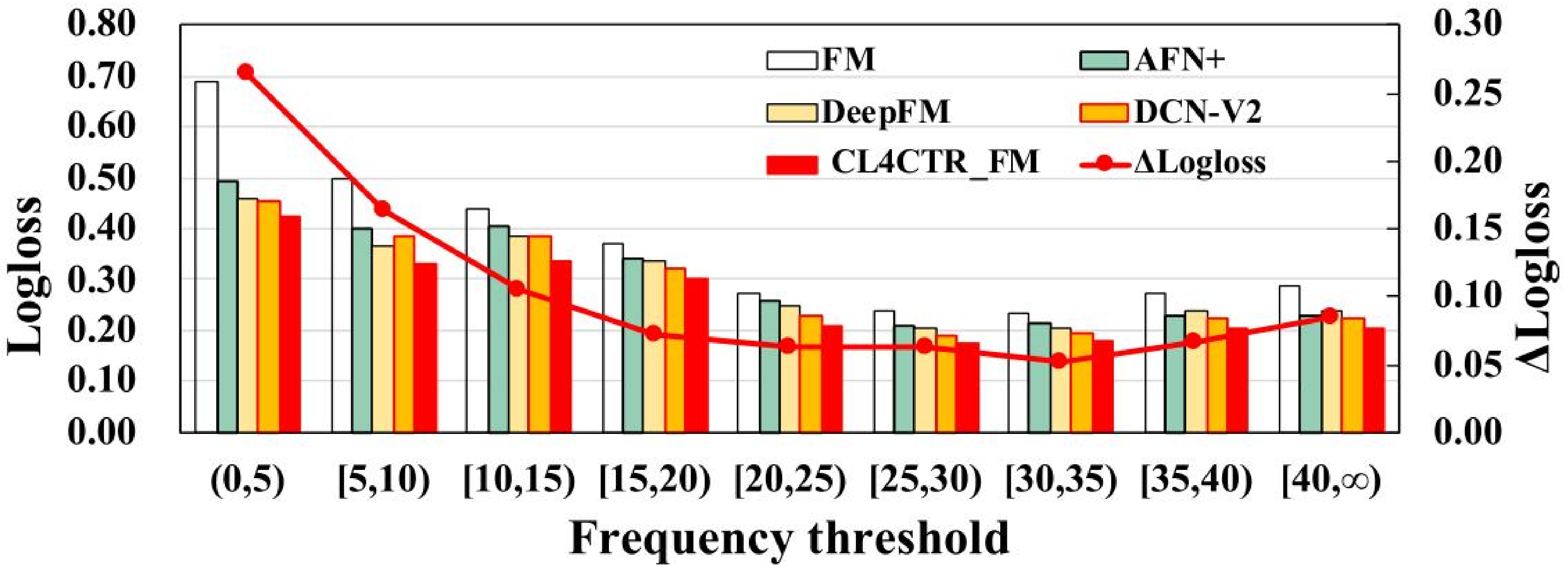
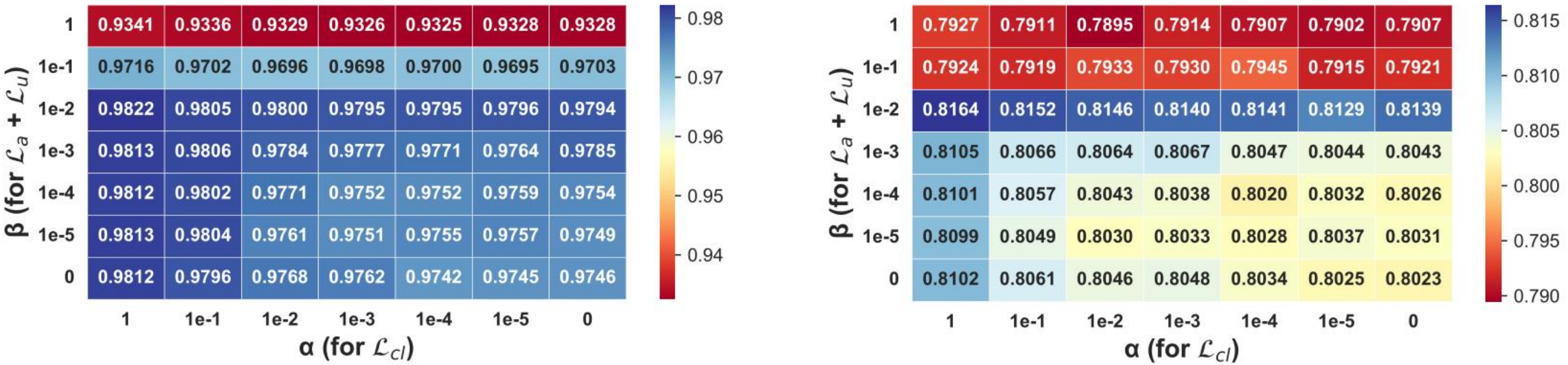


Figure 3: Improvement vs. feature frequency.

Experiments



(a) The AUC of Frappe　　　(b) The AUC of ML-1M

Figure 4: Performance of $CL4CTR_{FM}$ w.r.t. different weights assigned to three SSL signals: α for $\mathcal{L}_c$, β for $\mathcal{L}_a$ and $\mathcal{L}_u$.

Experiments

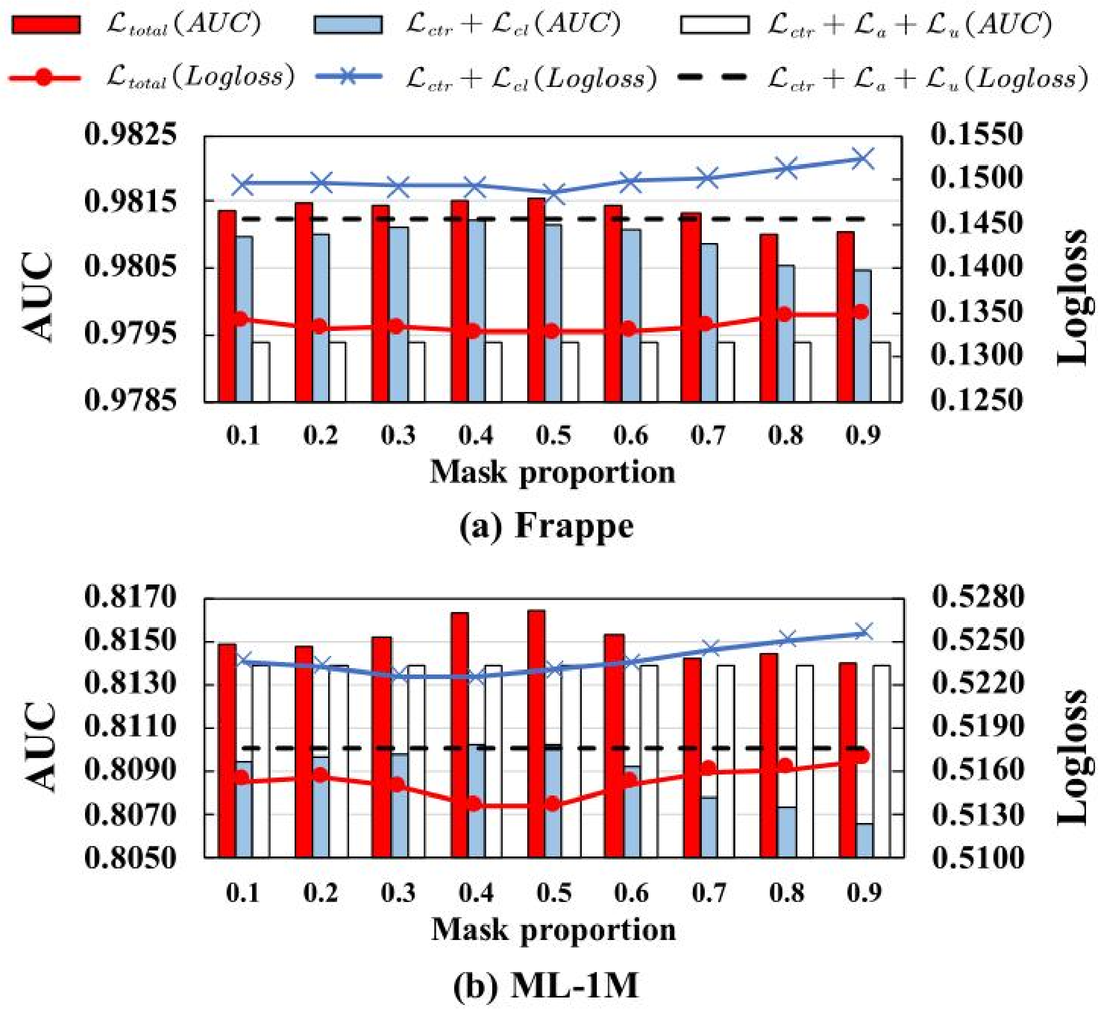


Figure 5: Impact of random mask proportion.

Experiments

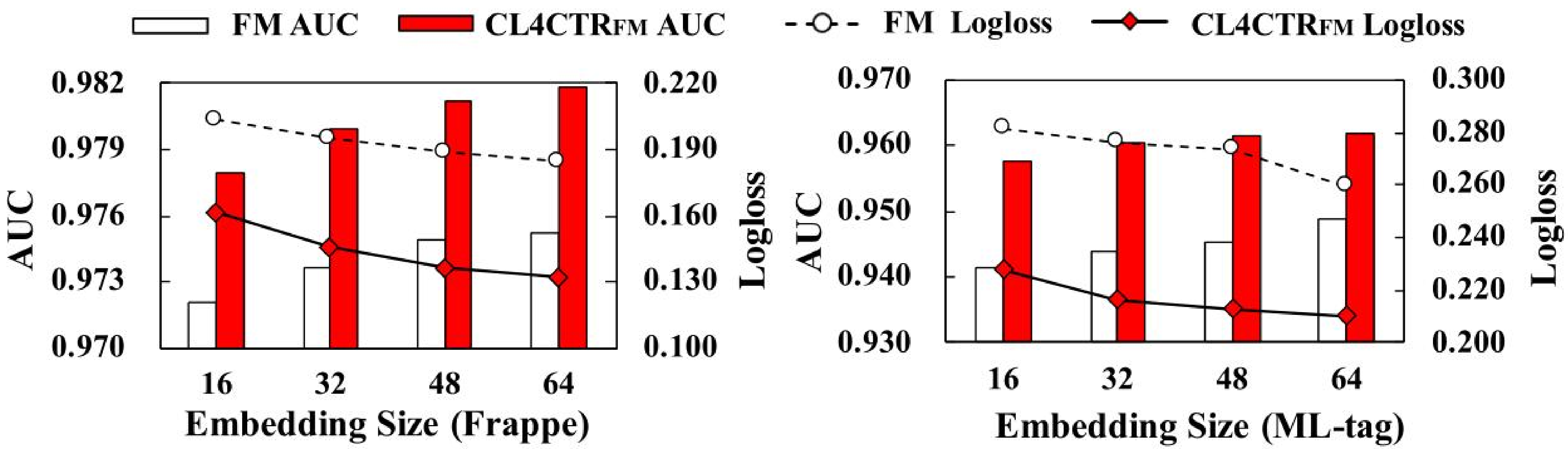


Figure 6: Impact of embedding size on FM and $CL4CTR_{FM}$.

Thanks