On the Embedding Collapse When Scaling Up Recommendation Models

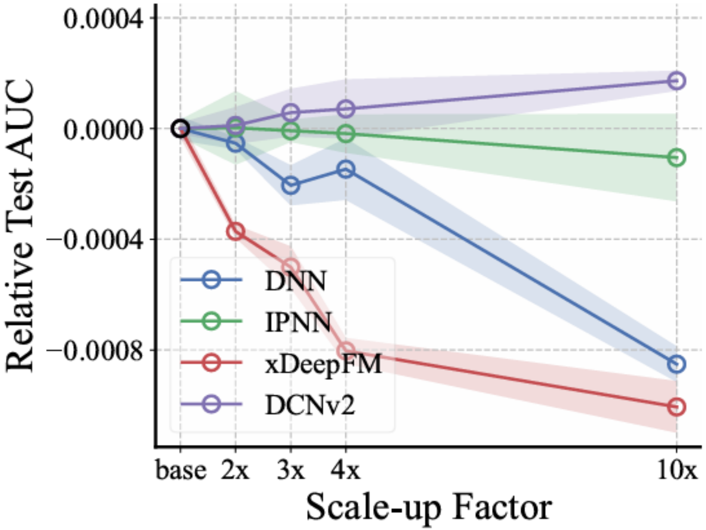
Xingzhuo Guo Junwei Pan Ximei Wang Baixu Chen Jie Jiang Mingsheng Long

Gustavo Cornejo Juanita Fernández Francisco Jorquera

Contexto

Problema

Mala escalabilidad de los modelos de recomendación existentes



(a) Performance when scaling up recommendation models

Trabajos relacionados

Módulos de RecSys

- Propuestas de diversos modelos
- No se estudia escalabilidad

Fenómeno del colapso

- Estudios del fenómeno para machine learning
- Falta de estudios para sistemas recomendadores

Teoría de Compresión

- Teorías para describir la complejidad de los datos

Contribución

Model Scalability issue

Embedding collapse

Two-sided effect

Feature interaction

Simple unified design

Multi-embedding design

Sistemas Recomendadores

$$\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \ldots \times \mathcal{X}_N$$
$$\mathcal{X}_i = \{1, 2, \ldots, D_i\}$$

$$\rightarrow \quad \mathcal{Y} = \{0, 1\}$$

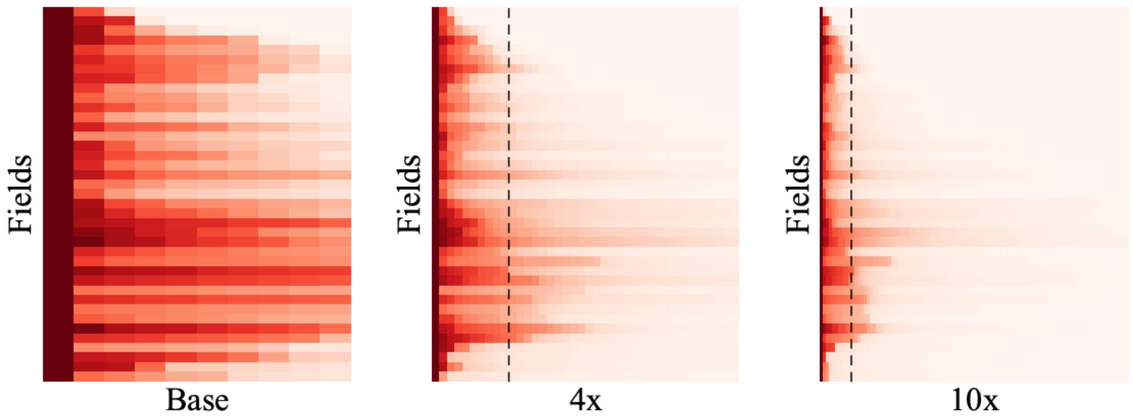
$$\boldsymbol{E}_i \in \mathbb{R}^{D_i \times K}$$

$$\boldsymbol{e}_i = \boldsymbol{E}_i^\top \mathbf{1}_{x_i}, \; \forall i \in \{1, 2, \ldots, N\},$$
$$h = I(\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n),$$
$$\hat{y} = F(h),$$

Embedding Collapse

Embedding Collapse

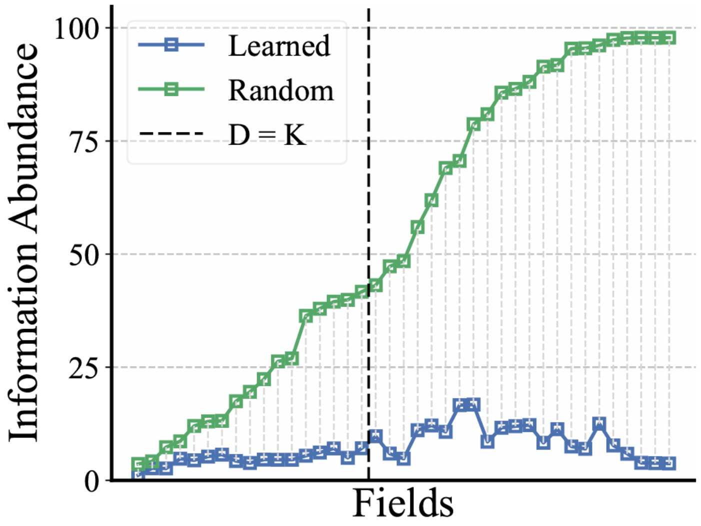
Matrices del embedding son de rango bajo



(b) Singular values of DCNv2 under different model size, with the dashed lines corresponding to the base size.

Information Abundance

$$\mathrm{IA}(\boldsymbol{E}) = \frac{\|\boldsymbol{\sigma}\|_1}{\|\boldsymbol{\sigma}\|_\infty}$$



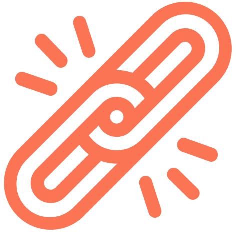
Feature Interaction

Feature Interaction

1. Embedding collapse

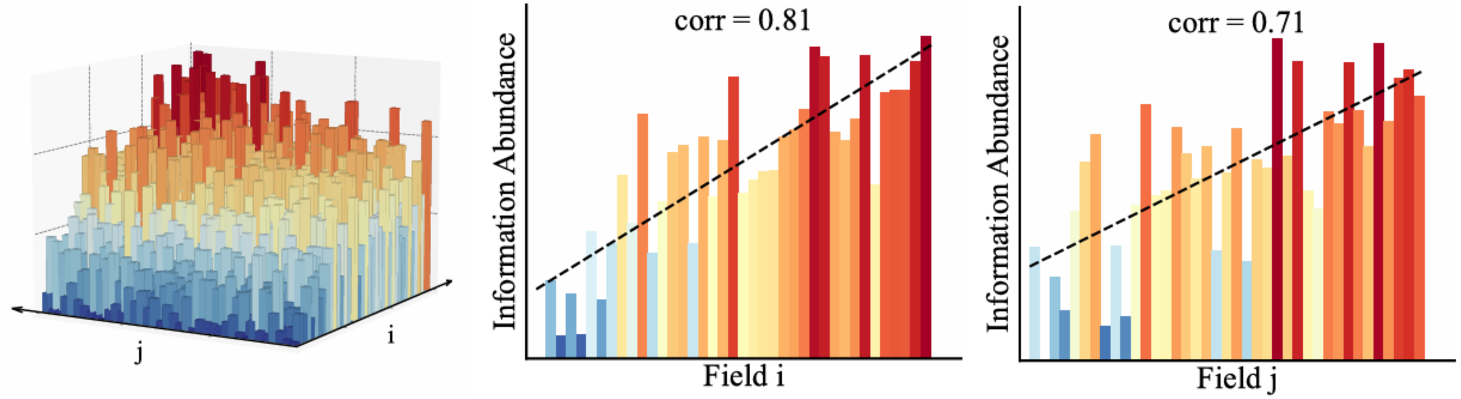
VS

2. Overfitting resistance



1. Interaction-Collapse Theory

Empirical analysis on models with sub embeddings



(a) $\mathrm{IA}(\boldsymbol{E}_i^{\to j})$. (b) $\sum_{j=1}^{N} \mathrm{IA}(\boldsymbol{E}_i^{\to j})$. (c) $\sum_{i=1}^{N} \mathrm{IA}(\boldsymbol{E}_i^{\to j})$.

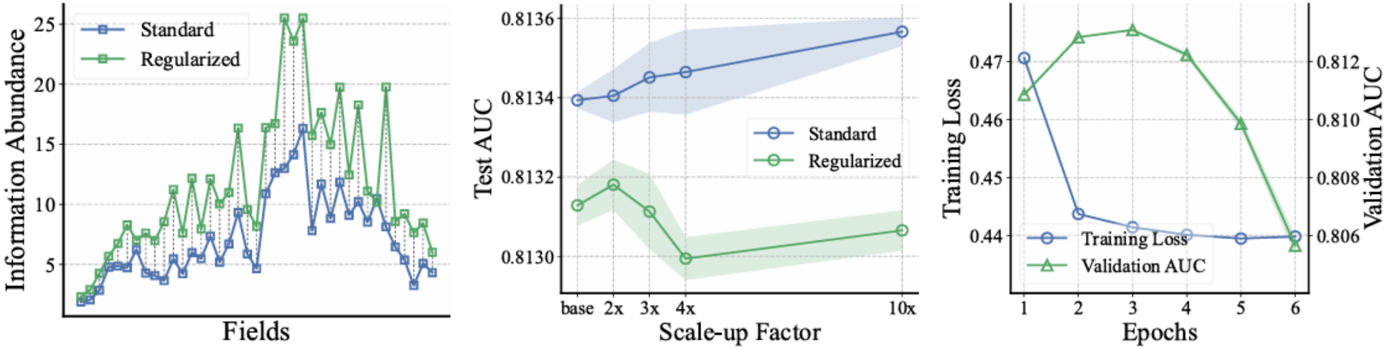
1. Interaction-Collapse Theory

How is embedding collapse caused?

Finding 1 (Interaction-Collapse Theory). In feature interaction of recommendation models, fields with low-information-abundance embeddings constrain the information abundance of other fields, resulting in collapsed embedding matrices.

2. Avoiding Collapse

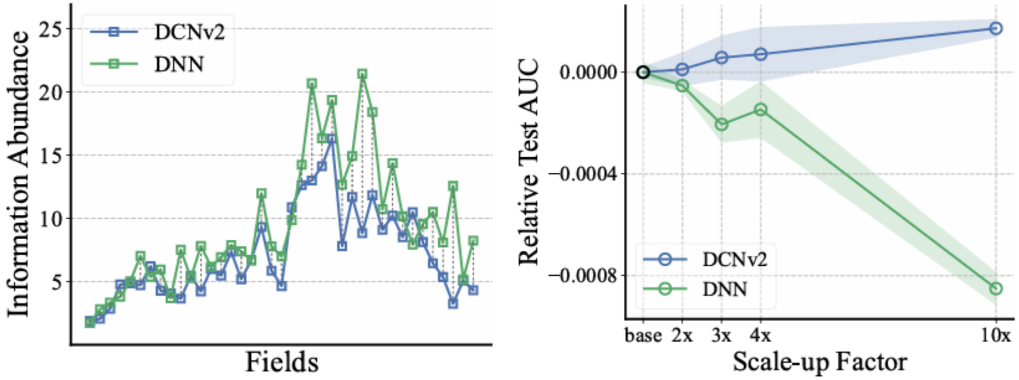
Limiting the modules in interaction that leads to collapse



(a) IA w/ 10x size. (b) Test AUC w.r.t. size. (c) Training curve.

2. Avoiding Collapse

Directly avoiding explicit interaction



(a) IA w/ 10x size. (b) Test AUC w.r.t. size.

2. Avoiding Collapse

Finding 2. A less-collapsed model with feature interaction suppressed improperly is insufficient for scalability due to overfitting concern.

Multi-Embedding

Multi-Embedding

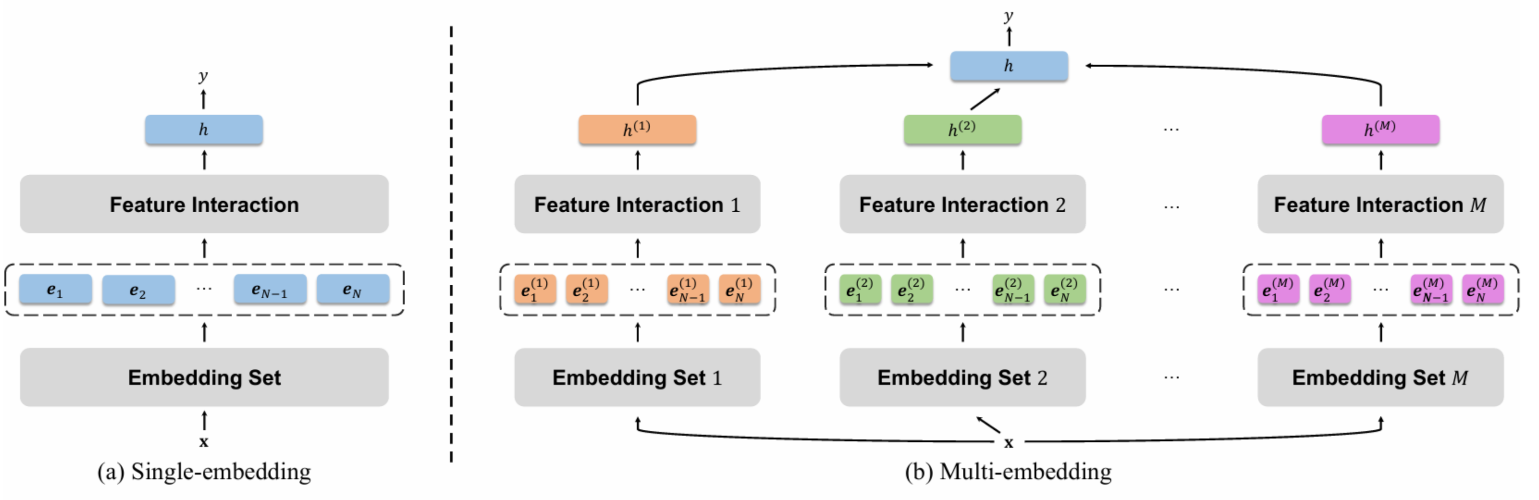
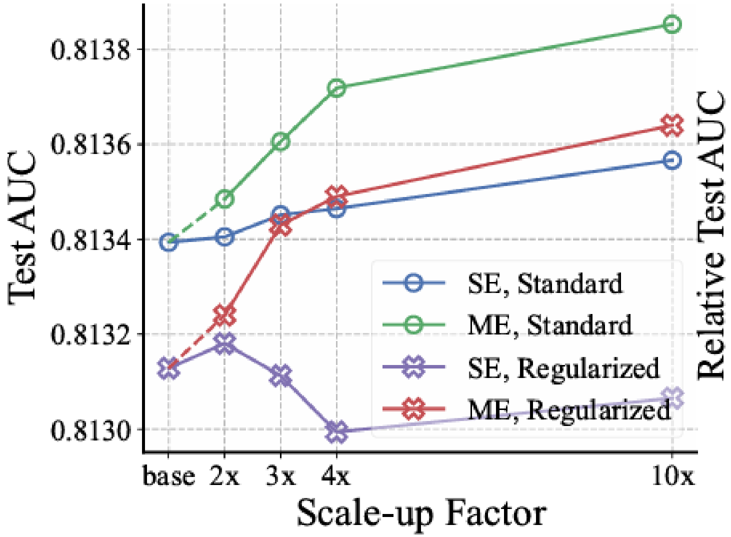


Figure 7. Architectures of single-embedding (left) and multi-embedding (right) models.

Resultados

Model		Criteo base	Criteo 2x	Criteo 3x	Criteo 4x	Criteo 10x	Avazu base	Avazu 2x	Avazu 3x	Avazu 4x	Avazu 10x
DNN	SE	<u>0.81228</u>	<u>0.81222</u>	0.81207	0.81213	0.81142	0.78744	<u>0.78759</u>	0.78752	0.78728	0.78648
	ME		0.81261	**0.81288**	**0.81289**	**0.81287**		0.78805	0.78826	0.78862	**0.78884**
IPNN	SE	<u>0.81272</u>	<u>0.81273</u>	<u>0.81272</u>	<u>0.81271</u>	0.81262	0.78732	<u>0.78741</u>	0.78738	<u>0.78750</u>	<u>0.78745</u>
	ME		0.81268	0.81270	0.81273	**0.81311**		0.78806	0.78868	0.78902	**0.78949**
NFwFM	SE	0.81059	0.81087	0.81090	<u>0.81112</u>	<u>0.81113</u>	0.78684	0.78757	0.78783	<u>0.78794</u>	<u>0.78799</u>
	ME		0.81128	0.81153	0.81171	**0.81210**		0.78868	0.78901	0.78932	**0.78974**
xDeepFM	SE	<u>0.81217</u>	0.81180	0.81167	0.81137	0.81116	<u>0.78743</u>	<u>0.78750</u>	0.78714	0.78735	0.78693
	ME		0.81236	0.81239	0.81255	**0.81299**		0.78848	0.78886	0.78894	**0.78927**
DCNv2	SE	0.81339	0.81341	0.81345	0.81346	<u>0.81357</u>	0.78786	0.78835	<u>0.78854</u>	<u>0.78852</u>	<u>0.78856</u>
	ME		0.81348	0.81361	**0.81382**	**0.81385**		0.78862	0.78882	0.78907	**0.78942**
FinalMLP	SE	<u>0.81259</u>	<u>0.81262</u>	0.81248	0.81240	0.81175	0.78751	<u>0.78797</u>	<u>0.78795</u>	0.78742	0.78662
	ME		0.81290	**0.81302**	**0.81303**	**0.81303**		0.78821	**0.78831**	**0.78836**	**0.78830**

Resultados



(a) Standard vs Regularized

Conclusión

El diseño Multi-Embedding mejora la escalabilidad del modelo y reduce el colapso

Referencias

Jean-Baptiste Tien, joycenv, O. C. play advertising challenge, 2014. https://kaggle.com/competitions/criteo-display-ad-challenge.

Jing, L., Vincent, P., LeCun, Y., and Tian, Y. Understanding dimensional collapse in contrastive self-supervised learning. In ICLR, 2021

Rendle, S., Krichene, W., Zhang, L., and Anderson, J. Neural collaborative filtering vs. matrix factorization revisited. In RecSys, 2020.

Steve Wang, W. C. Click-through rate prediction, 2014. URL https://kaggle.com/competitions/avazu-ctr-prediction.

Wang, R., Shivanna, R., Cheng, D., Jain, S., Lin, D., Hong, L., and Chi, E. DCN V2: Improved Deep & Cross Net work and Practical Lessons for Web-scale Learning to Rank Systems. In WWW, 2021.

On the Embedding Collapse When Scaling Up Recommendation Models

Xingzhuo Guo Junwei Pan Ximei Wang Baixu Chen Jie Jiang Mingsheng Long

Gustavo Cornejo Juanita Fernández Francisco Jorquera