eXtreme Deep Factorization Machine (xDeepFM)

For Recommender Systems

Zhang Yu, Liu Mengling, Zhang Yu
Group Y | Nov 14, 2020

Background & Introduction

Traditional cross feature engineering

Task-specific
Web-scale
Unseen interactions

Factorization Machines

Including all features both useful and useless combinations which introduces noises

FNN & PNN

They focus more on high-order feature interactions while capture little low-order interactions

Background & Introduction

Wide & Deep & DeepFM

DNNs model high-order feature interactions in an implicit fashion at the bit-wise level

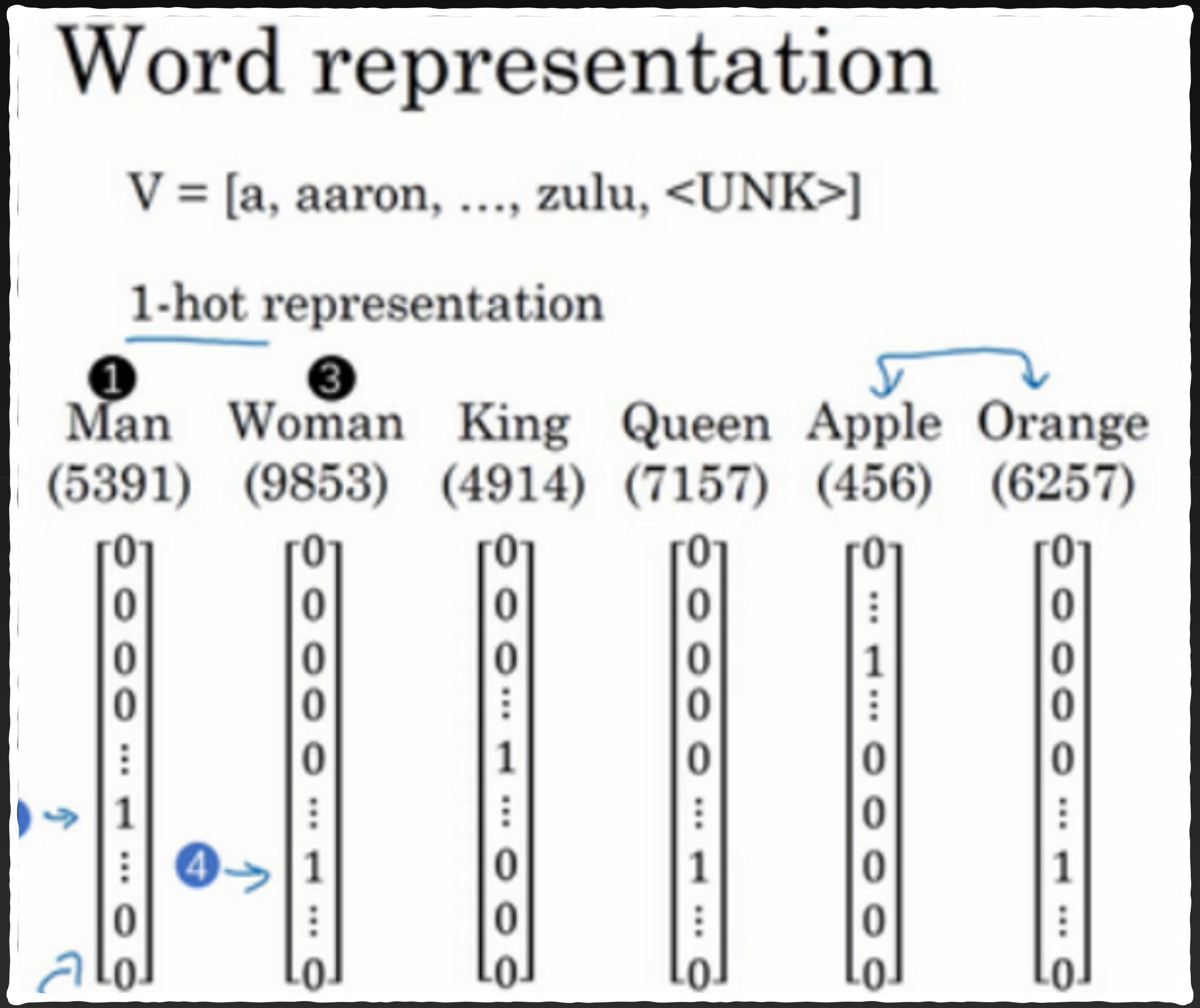
Cross Network (CrossNet)

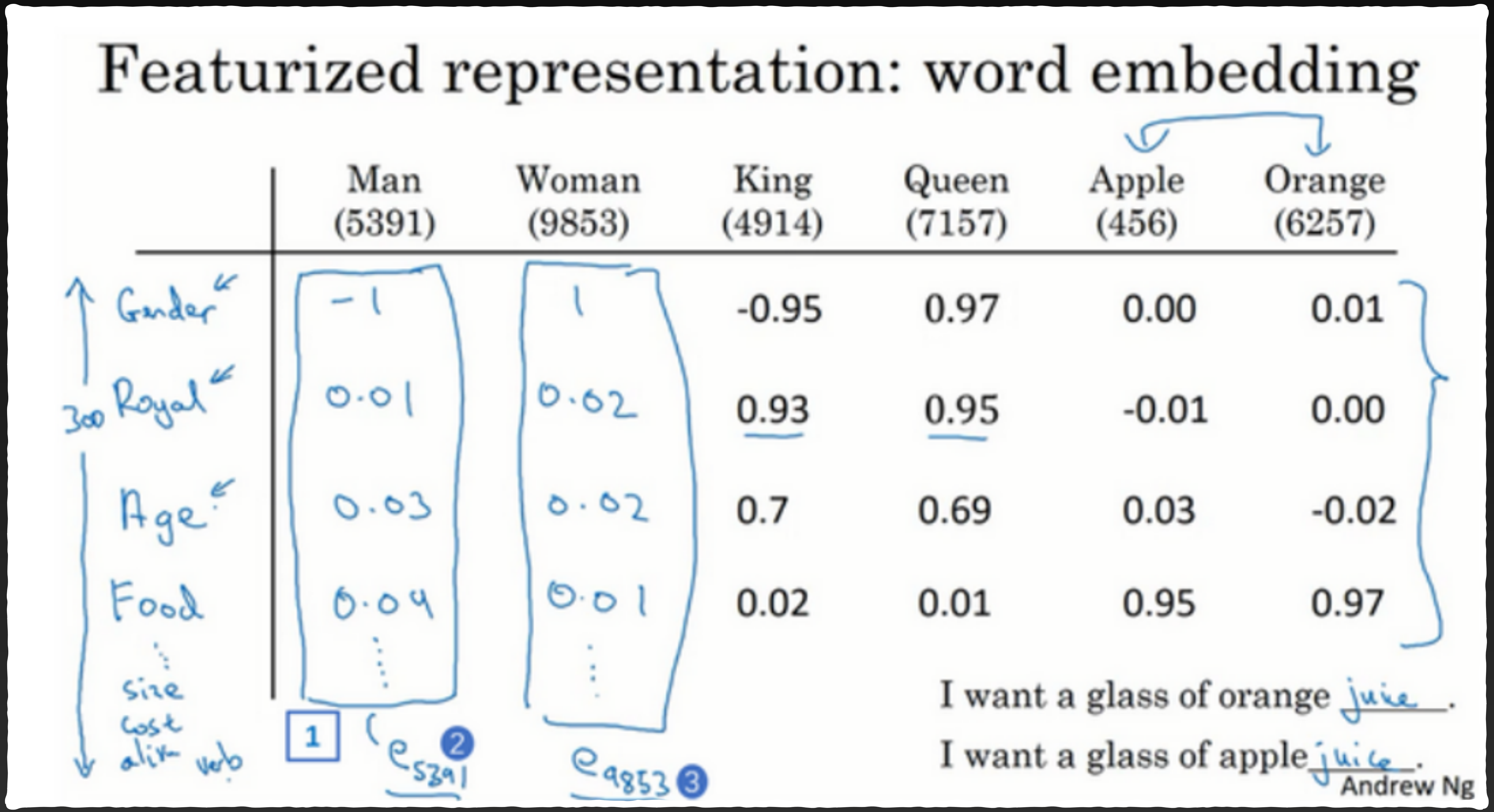
Each hidden layer is a scalar multiple of embedding layer; Interactions come in a bit-wise fashion

This paper propose a novel Compressed Interaction Network (CIN), which aims to generate feature interactions in an explicit fashion and at the vector-wise level.

Goal

Word Embedding





Reference: https://www.coursera.org/specializations/deep-learning?

Implicit and Explicit High-order Interactions

What is Implicit High-order Interactions?
It use feed-forward neural network on the field embedding vector to learn high-order feature interactions. It can not figure out the relationship between hidden layers and embedding layer.

$$\mathbf{x}^1 = \sigma(\mathbf{W}^{(1)}\mathbf{e} + \mathbf{b}^1) \quad (1)$$

$$\mathbf{x}^k = \sigma(\mathbf{W}^{(k)}\mathbf{x}^{(k-1)} + \mathbf{b}^k) \quad (2)$$

All formulas come from reference: Lian, Jianxun, et al. "xdeepfm: Combining explicit and implicit feature interactions for recommender systems." Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018.

What is Explicit High-order Interactions?
It shows that each hidden layer is a scalar multiple of the embedding layer which means that the relationship between hiden layers and embedding layer can be figured out.

$$\mathbf{x}_k = \mathbf{x}_0\mathbf{x}_{k-1}^T\mathbf{w}_k + \mathbf{b}_k + \mathbf{x}_{k-1} \quad (3)$$

$$\begin{aligned}\mathbf{x}_1 &= \mathbf{x}_0(\mathbf{x}_0^T\mathbf{w}_1) + \mathbf{x}_0 \\ &= \mathbf{x}_0(\mathbf{x}_0^T\mathbf{w}_1 + 1) \\ &= \alpha^1\mathbf{x}_0\end{aligned} \quad (4)$$

$$\begin{aligned}\mathbf{x}_{i+1} &= \mathbf{x}_0\mathbf{x}_i^T\mathbf{w}_{i+1} + \mathbf{x}_i \\ &= \mathbf{x}_0((\alpha^i\mathbf{x}_0)^T\mathbf{w}_{i+1}) + \alpha^i\mathbf{x}_0 \\ &= \alpha^{i+1}\mathbf{x}_0\end{aligned} \quad (5)$$

Vector-wise and Bit-wise Level

What is bit-wise level ?

$X_0 \in \mathbb{R}^{1\times(m*D)}$, it is means that X_0 only has one vector. Even the bits in the same field can be interacted with each other.

What is vector-wise level ?

It $X_0 \in \mathbb{R}^{m\times D}$, m is the number of features which means that it has m vectors.

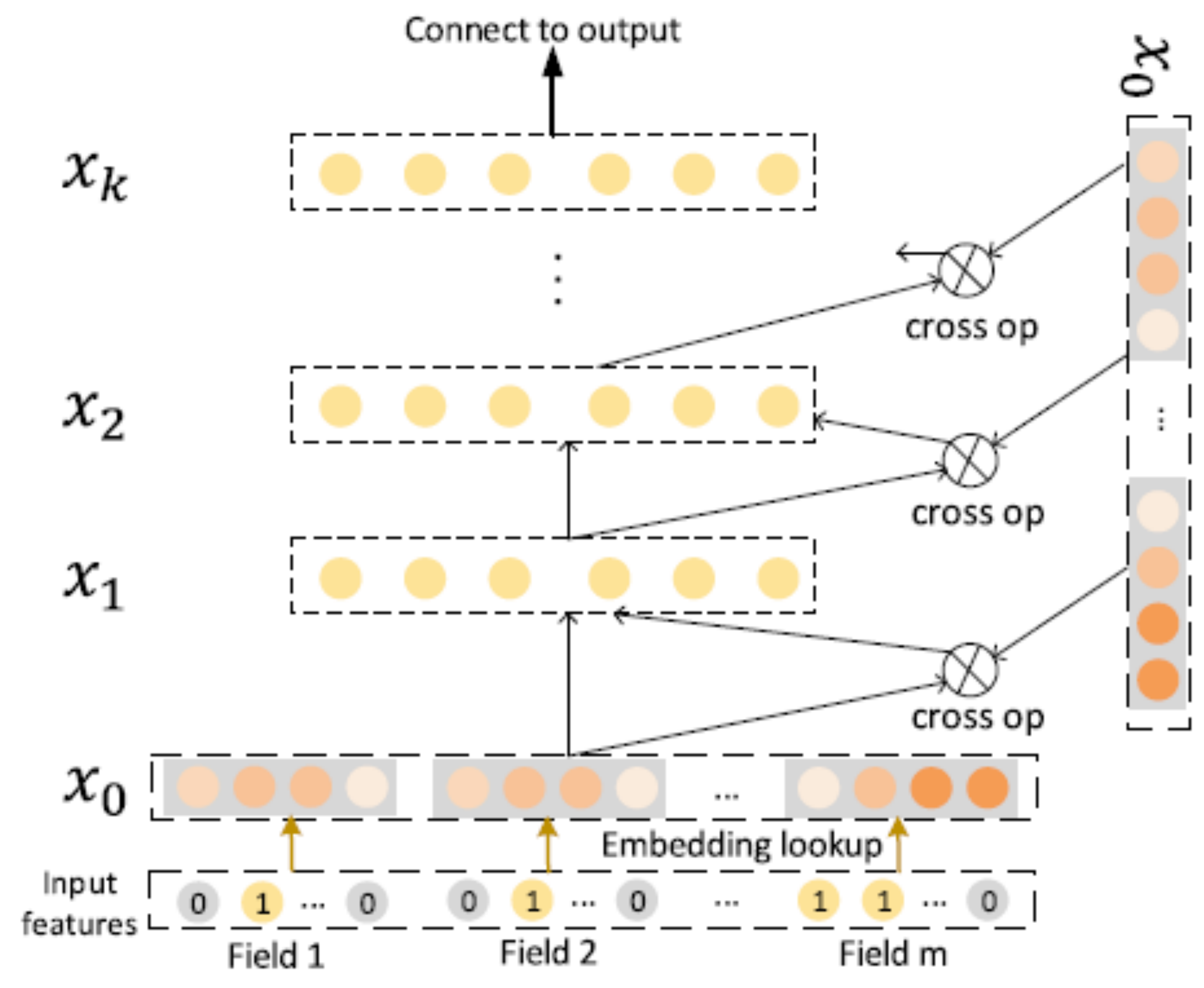
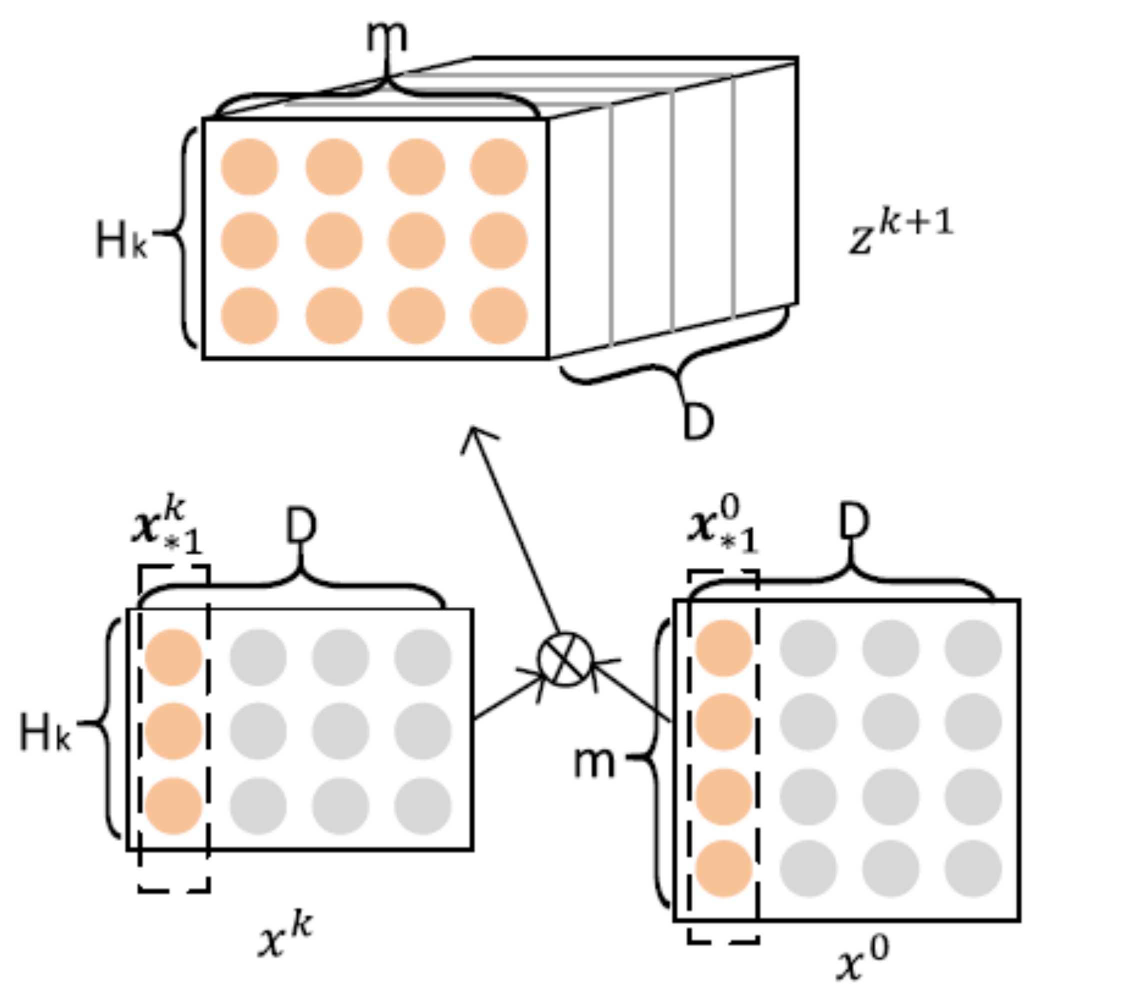


Figure 3: The architecture of the Cross Network.



(a) Outer products along each dimension for feature interactions. The tensor Z^{k+1} is an intermediate result for further learning.

All pictures come from reference: Lian, Jianxun, et al. "xdeepfm: Combining explicit and implicit feature interactions for recommender systems." Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018.

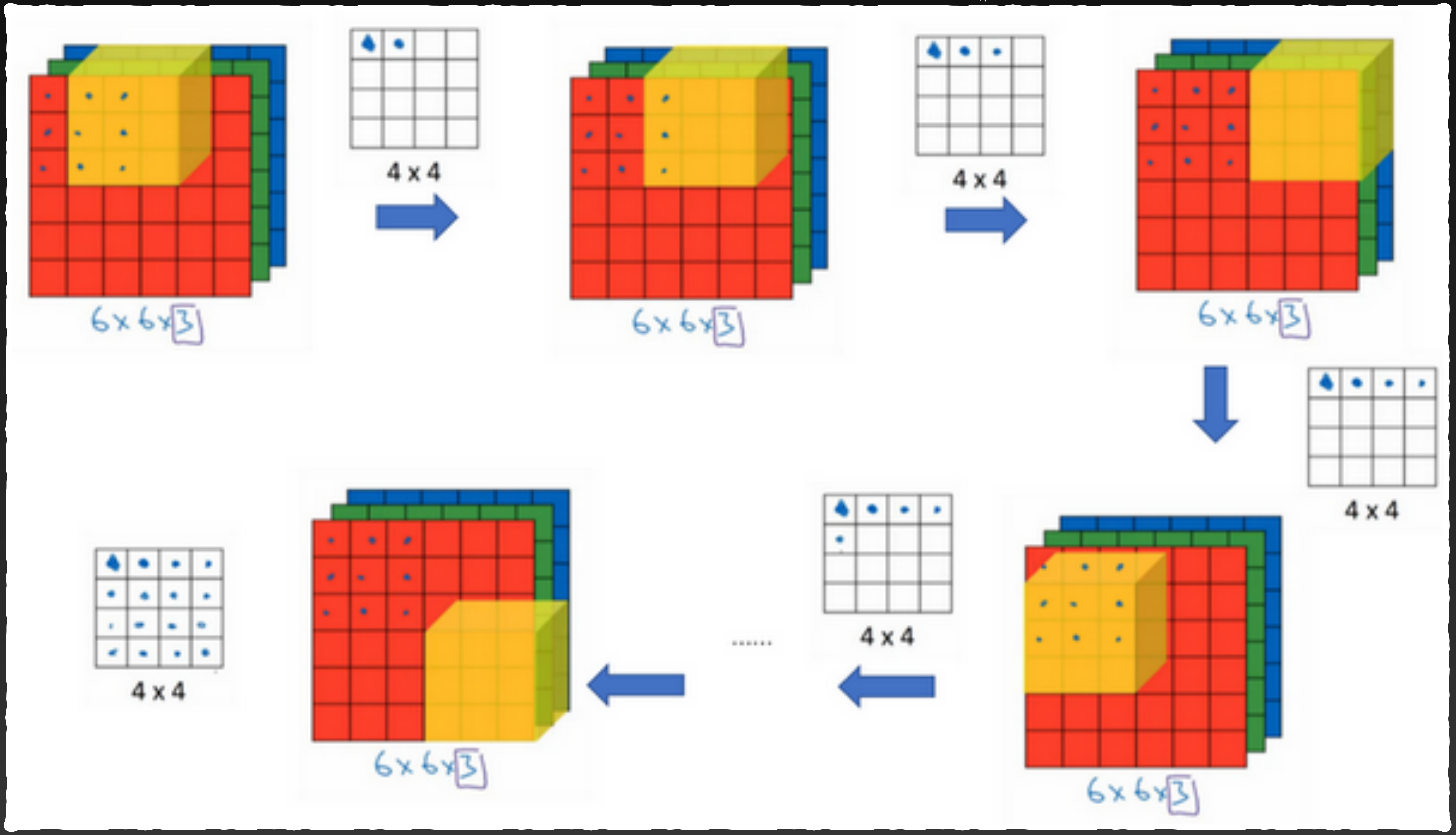
Filters in CNN

Intuitively: The data matrixes interacts with the filters with some rules.

Filters

Output

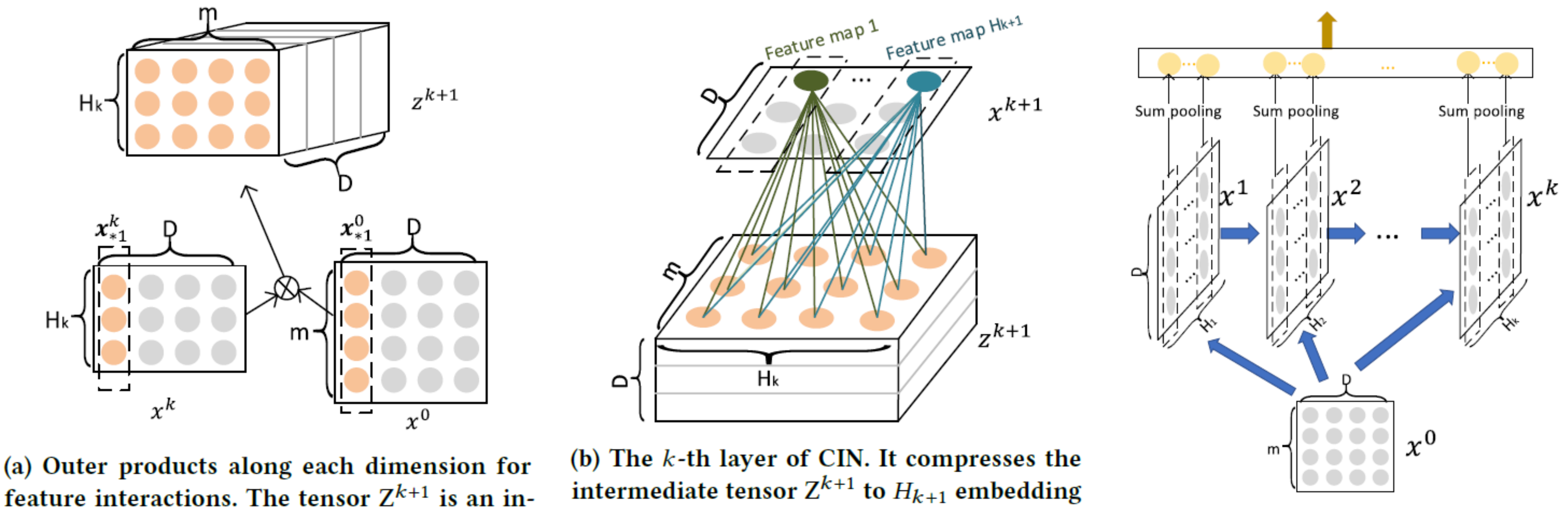
Data matrixes



Reference: https://www.coursera.org/specializations/deep-learning?

Compressed Interaction Network

Intuitively: It works same as the filters in the CNN. Using some filters to extract some features from the big data matrixes



(a) Outer products along each dimension for feature interactions. The tensor Z^{k+1} is an intermediate result for further learning.

(b) The k-th layer of CIN. It compresses the intermediate tensor Z^{k+1} to H_{k+1} embedding vectors (aslo known as *feature maps*).

(c) An overview of the CIN architecture.

All pictures come from reference: Lian, Jianxun, et al. "xdeepfm: Combining explicit and implicit feature interactions for recommender systems." Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018.

xDeepFM

Combining Explicit and Implicit Feature Interactions

Explicit component

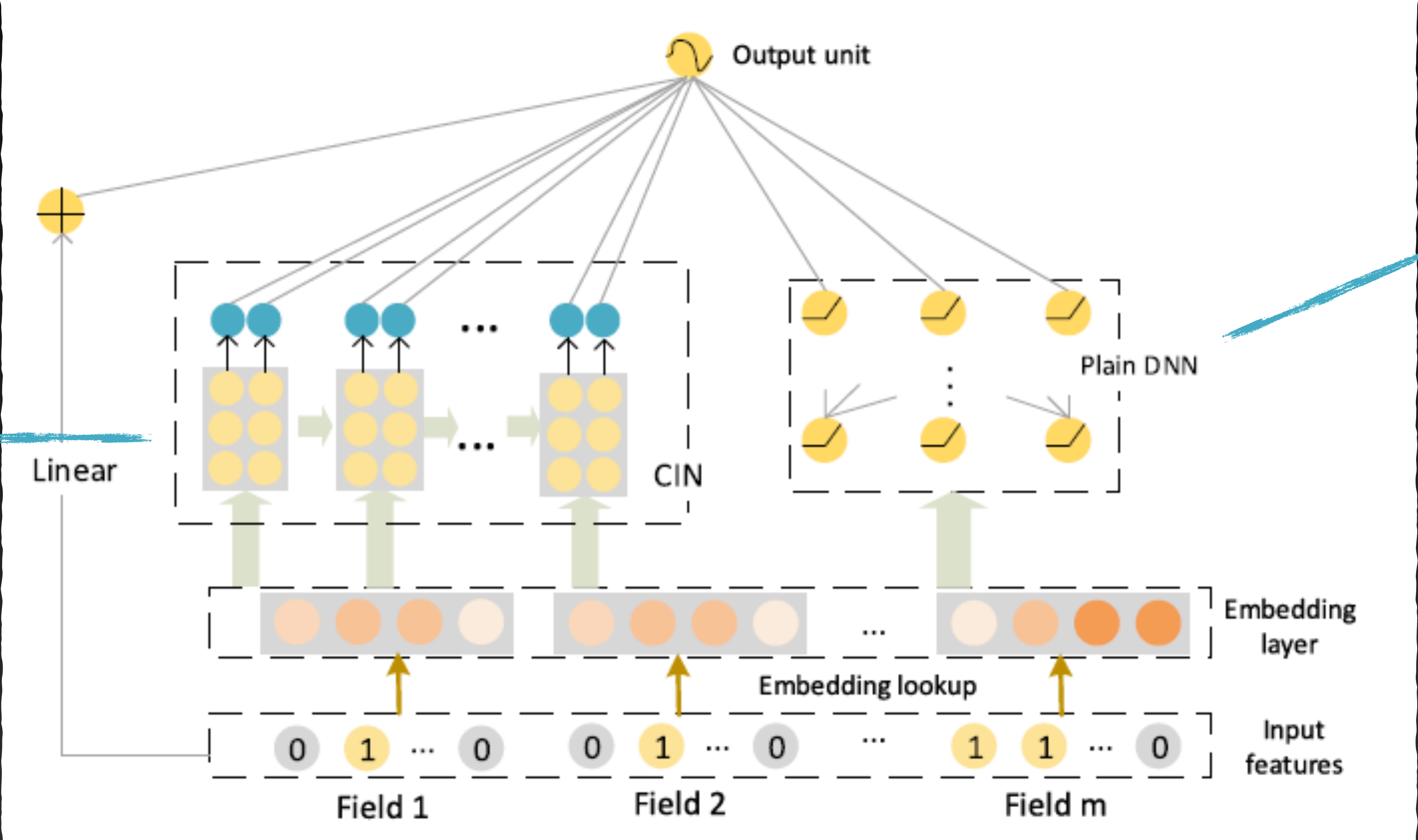


Figure 5: The architecture of xDeepFM.

Implicit component

All pictures come from reference: Lian, Jianxun, et al. "xdeepfm: Combining explicit and implicit feature interactions for recommender systems." Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018.

Result

	Criteo			Dianping			Bing News		
Model name	AUC	Logloss	Depth	AUC	Logloss	Depth	AUC	Logloss	Depth
LR	0.7577	0.4854	-,-	0.8018	0.3608	-,-	0.7988	0.2950	-,-
FM	0.7900	0.4592	-,-	0.8165	0.3558	-,-	0.8223	0.2779	-,-
DNN	0.7993	0.4491	-,2	0.8318	0.3382	-,3	0.8366	0.2730	-,2
DCN	0.8026	0.4467	2,2	0.8391	0.3379	4,3	0.8379	0.2677	2,2
Wide&Deep	0.8000	0.4490	-,3	0.8361	0.3364	-,2	0.8377	0.2668	-,2
PNN	0.8038	0.4927	-,2	0.8445	0.3424	-,3	0.8321	0.2775	-,3
DeepFM	0.8025	0.4468	-,2	0.8481	0.3333	-,2	0.8376	0.2671	-,3
xDeepFM	**0.8052**	**0.4418**	3,2	**0.8639**	**0.3156**	3,3	**0.8400**	**0.2649**	3,2

All pictures come from reference: Lian, Jianxun, et al. "xdeepfm: Combining explicit and implicit feature interactions for recommender systems." Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018.

Thank you!

Zhang Yu, Liu Mengling, Zhang Yu
Group Y | Nov 14, 2020