Deep Factorization Machines

Đỗ Hoàng Thuấn

October 2019

Table of contents

Introduction

- Deep factorization machine is SOTA in Click Through Rate problem.
- DeepFM makes a good result on Zalo Hit Song Challenge.

Problem statement

Deep FM is build based on FM.

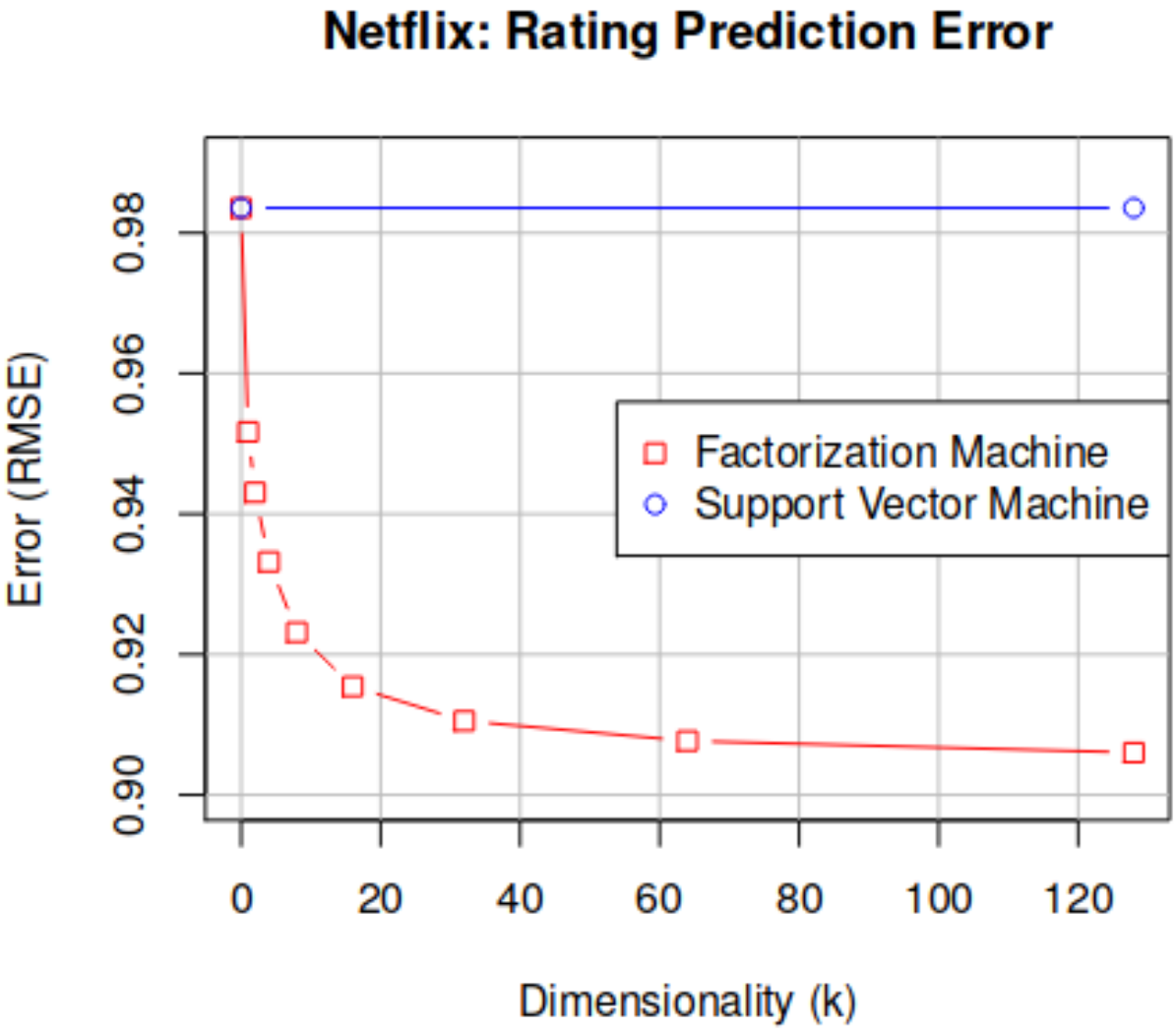
Suppose we have a dataset.

user						movie (item)					time	rating	
$x^{(i)}$	u_1	u_2	u_3	u_4	...	i_1	i_2	i_3	i_3	...	t	r	$y^{(i)}$
$x^{(1)}$	1	0	0	0	...	1	0	0	0	...	2	5	$y^{(1)}$
$x^{(2)}$	0	1	0	0	...	0	0	1	0	...	18	1	$y^{(2)}$
$x^{(3)}$	0	1	0	0	...	0	0	0	1	...	6	2	$y^{(3)}$
$x^{(4)}$	0	0	1	0	...	0	1	0	0	...	12	3	$y^{(4)}$
$x^{(5)}$	1	0	0	0	...	0	0	1	0	...	3	5	$y^{(5)}$
:	:	:	:	:	:	:	:	:	:	:	:	:	:
$x^{(m)}$	0	0	0	1	...	0	1	0	0	...	9	4	$y^{(m)}$

where:

- $X \in R^{m \times n}$
- m is number of observations, $n = n_{user} + n_{item} + n_T$
- x is feature vector, $v \int$

Result of FM


Netflix: Rating Prediction Error
Error (RMSE)
0.90
0.92
0.94
0.96
0.98
0
20
40
60
80
100
120
Dimensionality (k)
Factorization Machine
Support Vector Machine

Polynomial model and limitation

Apply polynomial(linear regresson).

$$\sigma = w_0 + \sum_{i=1}^{n} w_i x_i \tag{1}$$

where:

- w_0 is bias
- w_i is weight of field i
- x_i is feature vector of field i

Polynomial does not use combination of features.
Can not get the interaction between features.

Factorization Machine

Extends from logistic regression.
Equation of 2-order FM.

$$\hat{y}(x) = w_0 + \sum_{i=1}^{i=n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} < v_i, v_j > x_i x_j$$

Where:

w_0 is bias

w_i is weight of field i.

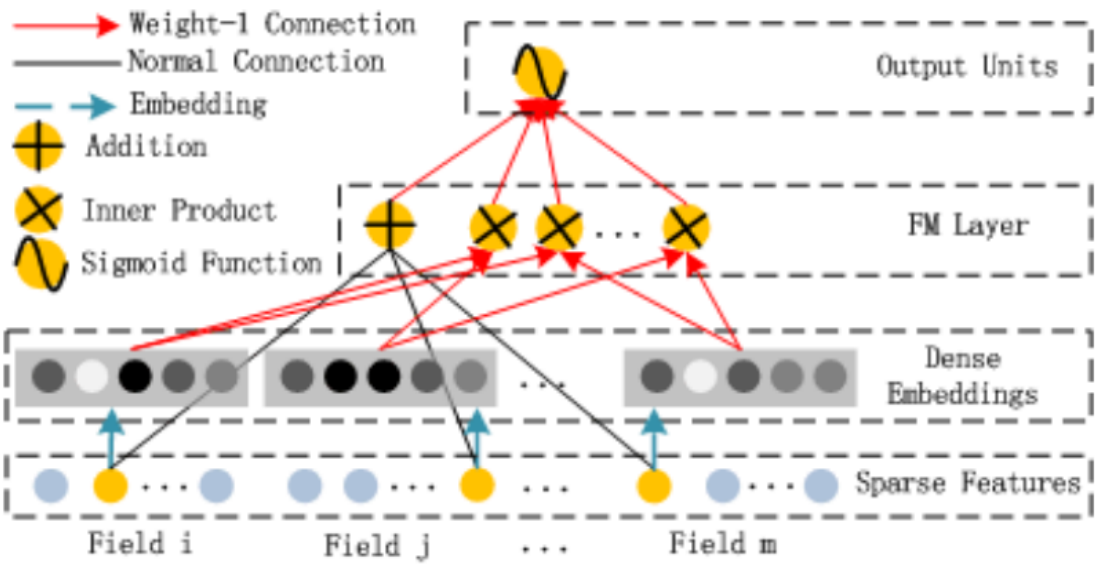
x_i is feature vector of field i.

v is latent feature vector.

$\hat{w}_{i,j} = n < v_i, v_j >$ is interaction between the i-th and j-th variable.

Factorization Machine Layer

Intergrate the architecture of FM and deep neural networks.

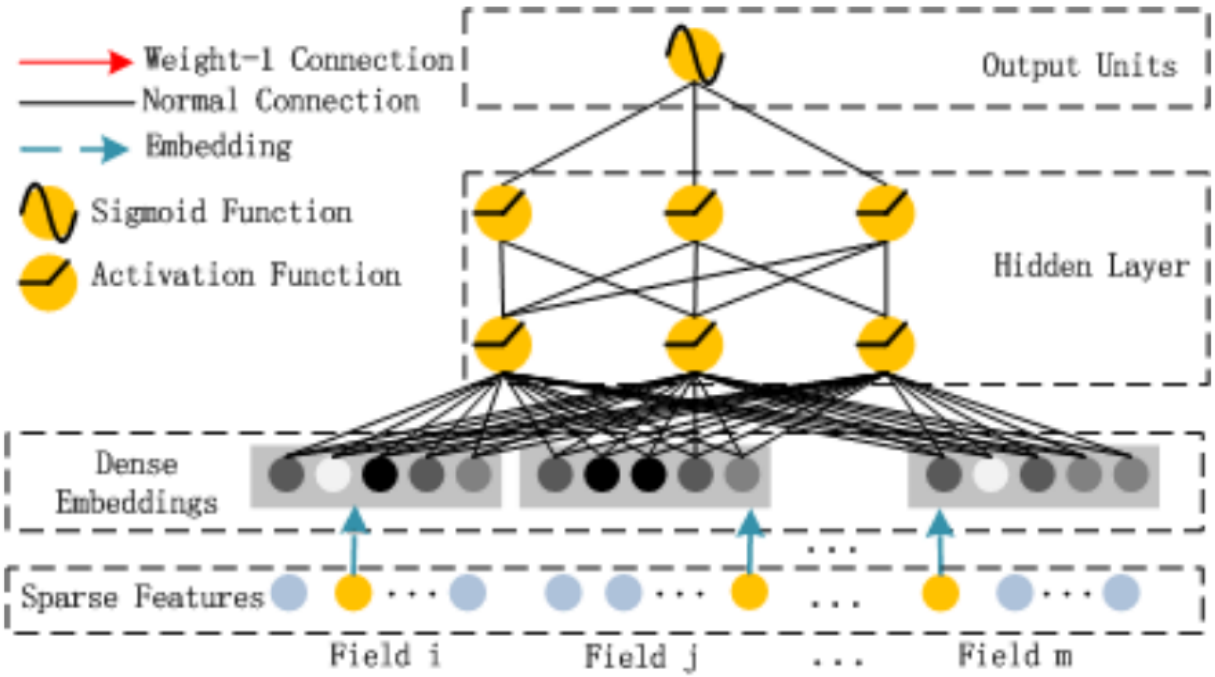


Hình: FM layer in DeepFM model

$$y_{out} = \sigma(w_0 + \sum_{i=1}^{i=n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} < v_i, v_j > x_i x_j)$$

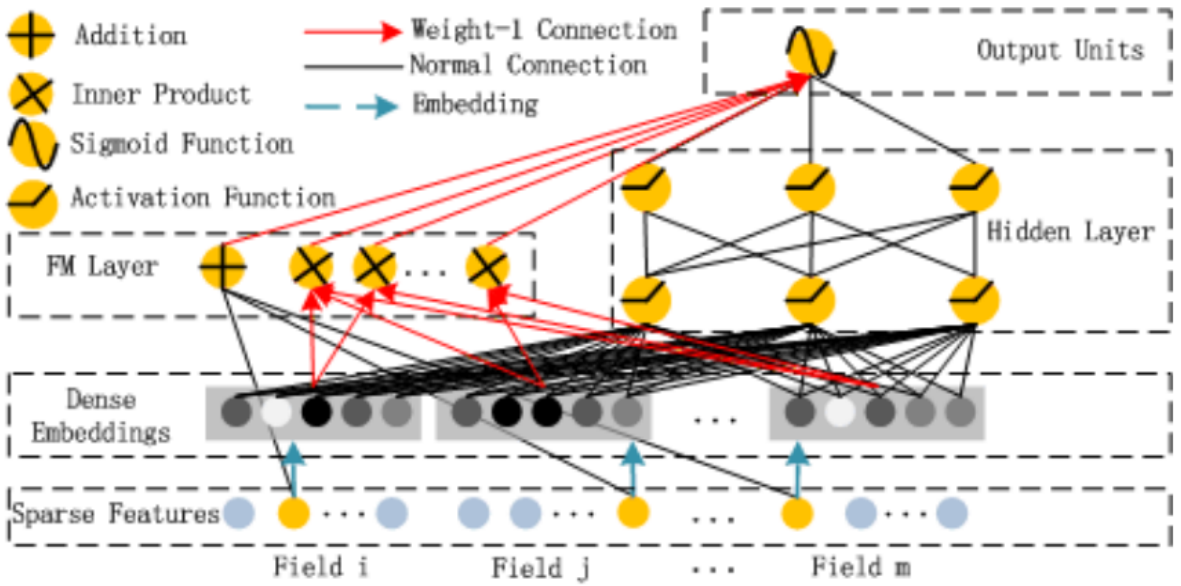
Deep component

Deep component is a feed-forward neural network.
It is used to learn high-order feature interactions.



Hình: DNN layer in DeepFM model

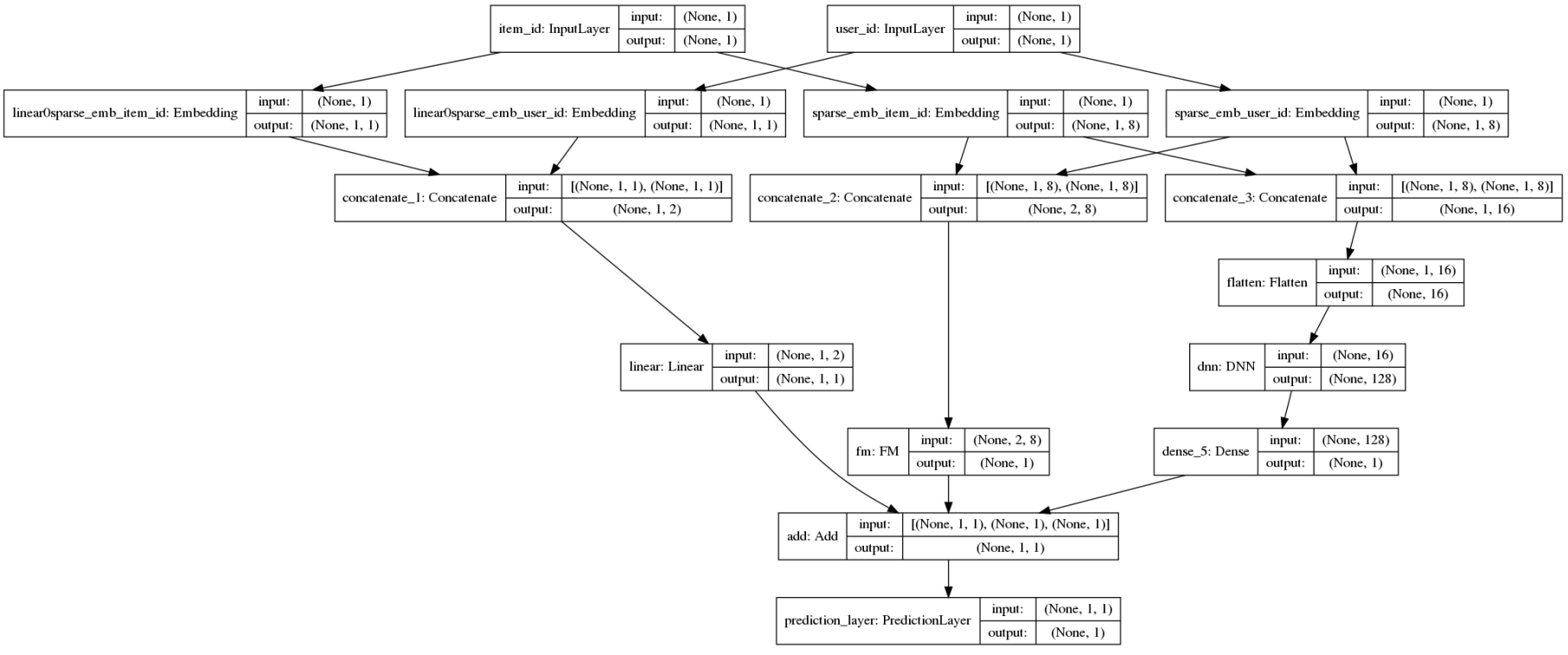
Full DeepFM architecture



Hình: Wide & deep architecture of DeepFM.

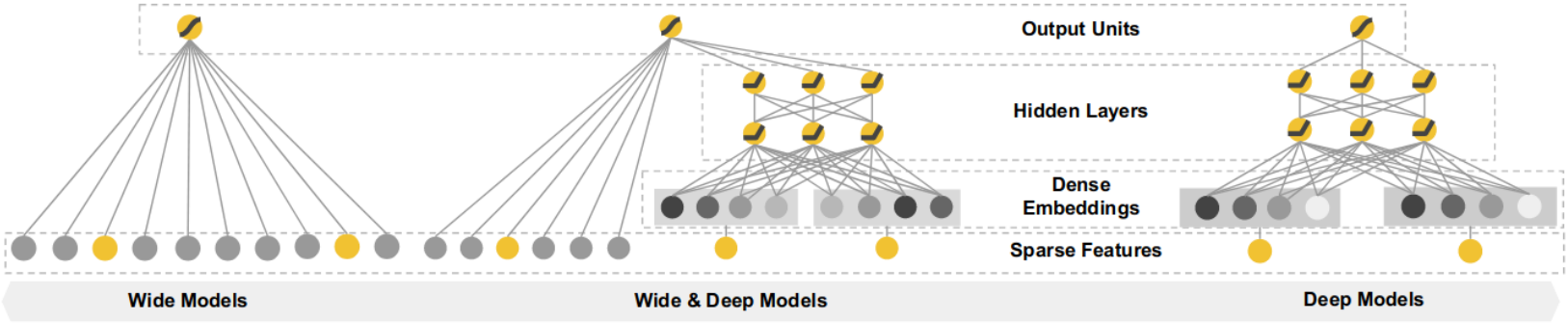
Need to add a part same as Wide& Deep's wide part.

Full DeepFM architecture



Hình: Wide & deep architecture of DeepFM.

Wide and deep model



Hình: Wide and Deep model

- W&D can model low and high feature interactions simultaneously.
- It needs for manumal expertise feature engineering on the input.
- W&D do not generalize to features pairs that have appeared in the training data.

Comparison

	Wide& Deep	DeepFM
High-order feature	✓	✓
Low-order feature	✓	✓
No feature engineering	✓	x

Experiments

Table 2: Performance on CTR prediction.

	Company∗		Criteo	
	AUC	LogLoss	AUC	LogLoss
LR	0.8640	0.02648	0.7686	0.47762
FM	0.8678	0.02633	0.7892	0.46077
FNN	0.8683	0.02629	0.7963	0.45738
IPNN	0.8664	0.02637	0.7972	0.45323
OPNN	0.8658	0.02641	0.7982	0.45256
PNN∗	0.8672	0.02636	0.7987	0.45214
LR & DNN	0.8673	0.02634	0.7981	0.46772
FM & DNN	0.8661	0.02640	0.7850	0.45382
DeepFM	**0.8715**	**0.02618**	**0.8007**	**0.45083**

References

- Factorization Machines & their application on huge datasets
- Factorization Machine
- Factorization Machine Paper
- Deep Factorization Machine Paper