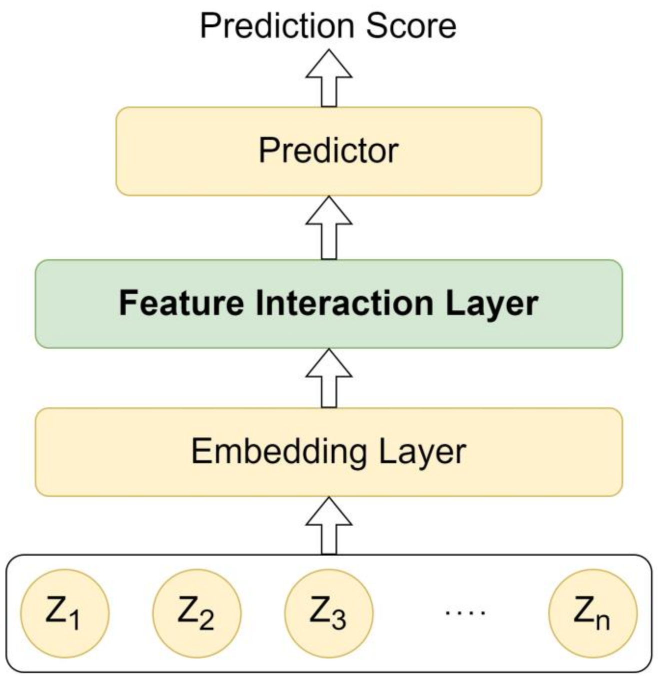
Towards Hybrid-grained Feature Interaction Selection for Deep Sparse Network

Fuyuan Lyu[1], Xing Tang[2*], Dugang Liu[4], Chen Ma[3],
Weihong Luo[2], Liang Chen[2], Xiuqiang He[2], Xue Liu[1]

[*]Corresponding Author
[1]School of Computer Science, McGill University
[2]FiT, Tencent; [3]City University of Hong Kong
[4]Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ)

Presenter: Fuyuan Lyu

Introduction

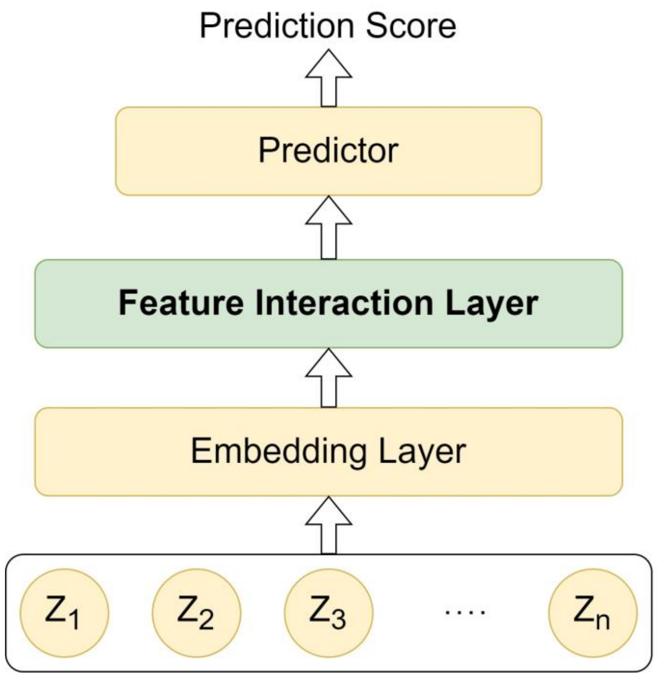


DSN framework

Deep Sparse Network: handle sparse (usually one-hot), high-dimensional features.

E.g. City, Season, Gender

Introduction



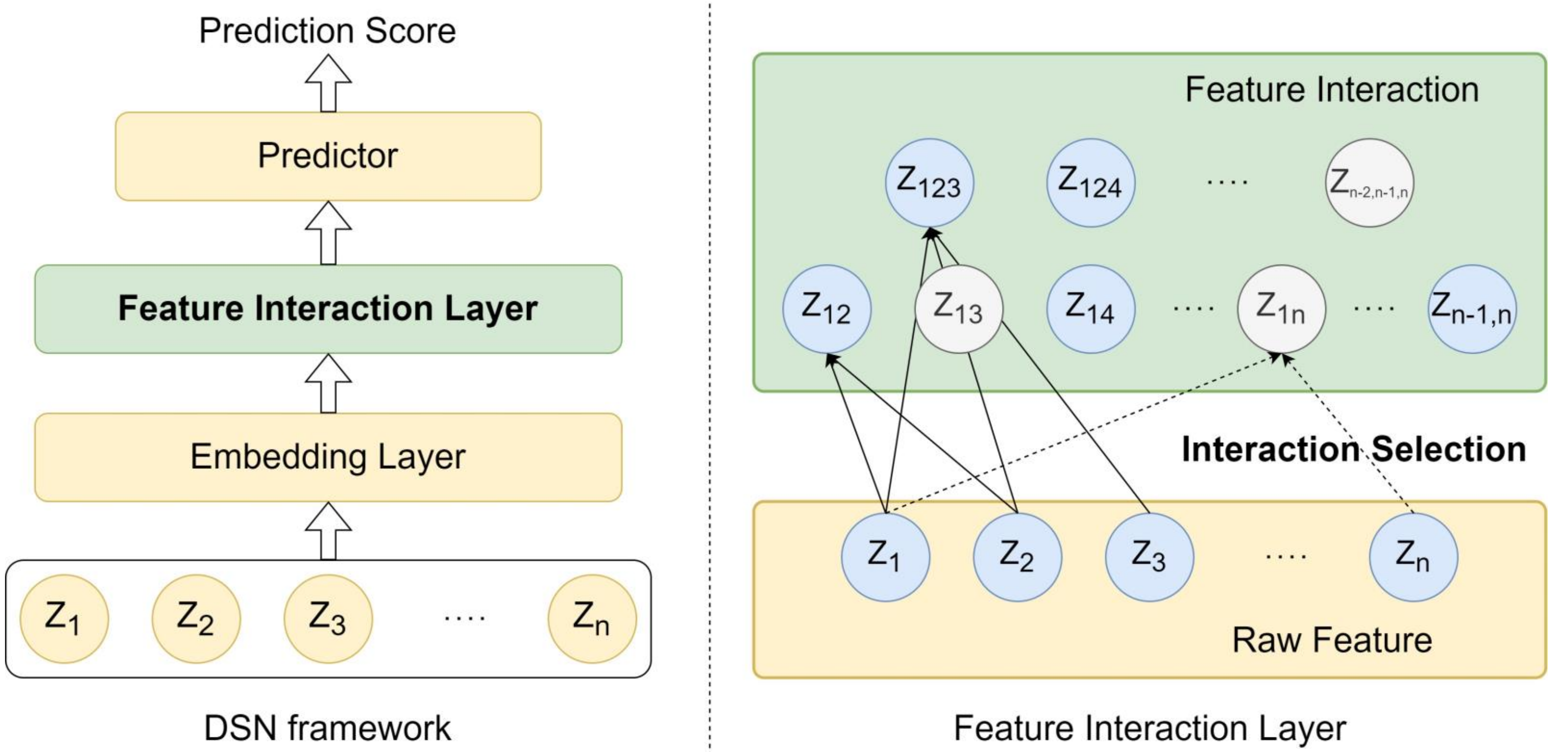
DSN framework

Deep Sparse Network: handle sparse (usually one-hot), high-dimensional features.

An example for feature interaction: <City, Season>

Feature Interaction Layer plays an important role in DSN framework

Introduction



Deep Sparse Network: handle sparse (usually one-hot), high-dimensional features.

An example for feature interaction: <City, Season>

Feature Interaction Layer plays an important role in DSN framework

Introduction

Consider yourself a sport advertiser

+

=

Montreal Winter Skiing!

Introduction

Consider yourself a sport advertiser

 +  = 

Montreal Winter Skiing!

Strong indicator.
Good!

Introduction

Consider yourself a sport advertiser

 +  = 

Montreal Winter Skiing!

Strong indicator.
Good!

 +  = 

Shanghai Winter Basketball?

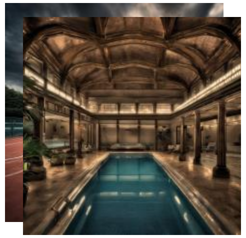
Introduction

Consider yourself a sport advertiser

 +  = 

Montreal Winter Skiing!

Strong indicator.
Good! 😀

 +  = 

Shanghai Winter Swimming?

Introduction

Consider yourself a sport advertiser

 +  = 

Montreal Winter Skiing!

Strong indicator.
Good!

 +  = 

Shanghai Winter Ping Pong?

Introduction

Consider yourself a sport advertiser

 +  = 

Montreal Winter Skiing!

Strong indicator. Good!

 +  = 

Shanghai Winter Ping Pong?

Perhaps not much info

Introduction

Consider yourself a sport advertiser

 +  = 

Montreal Winter Skiing!

Strong indicator. Good!

 +  = 

Shanghai Winter Ping Pong?

Perhaps not much info

Need a more fine-grained selection

Methodology

Dimension Explosion

For online advertisement systems

#Field $n \leq 100$

#Value $m \approx 10^7$

Methodology

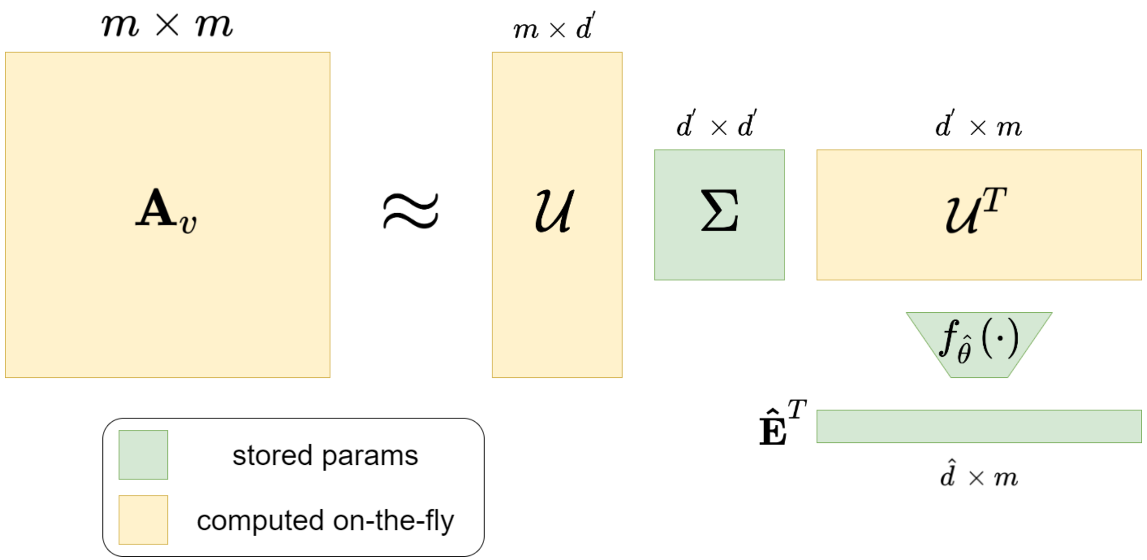
Dimension Explosion

For online advertisement systems

#Field $n \leq 100$

#Value $m \approx 10^7$

A is symmetrical.



Selection Tensor Decomposition

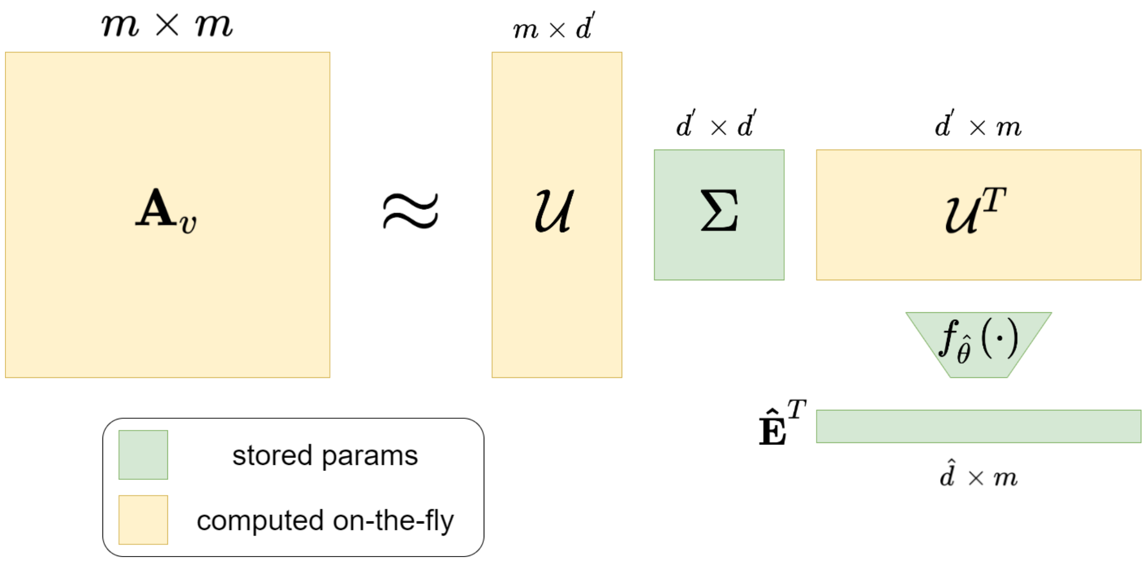
Methodology

Dimension Explosion

For online advertisement systems

#Field $n \le 100$

#Value $m \approx 10^7$



Selection Tensor Decomposition

$$\boldsymbol{A}^t = \alpha \boldsymbol{A}_f^t + (1 - \alpha)\boldsymbol{A}_v^t$$

$\boldsymbol{A}_f^t$ is the field-level selection
$\boldsymbol{A}_v^t$ is the value-level selection
$\alpha \in \{0, 1\}^{\{C_n^t\}}$ is the hybrid tensor

Hybrid-grained Selection

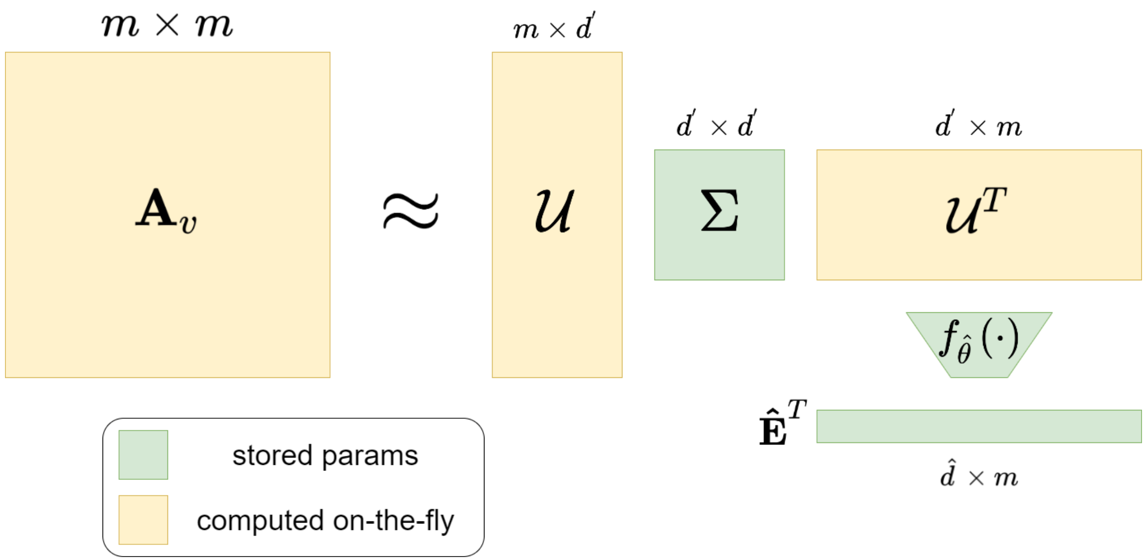
Methodology

Dimension Explosion

For online advertisement systems

#Field $n \leq 100$

#Value $m \approx 10^7$



Selection Tensor Decomposition

$$\boldsymbol{A}^t = \alpha \boldsymbol{A}_f^t + (1 - \alpha)\boldsymbol{A}_v^t$$

$\boldsymbol{A}_f^t$ is the field-level selection
$\boldsymbol{A}_v^t$ is the value-level selection
$\alpha \in \{0,1\}^{\{C_n^t\}}$ is the hybrid tensor

Hybrid-grained Selection

End-to-end selection

Selection results are discrete

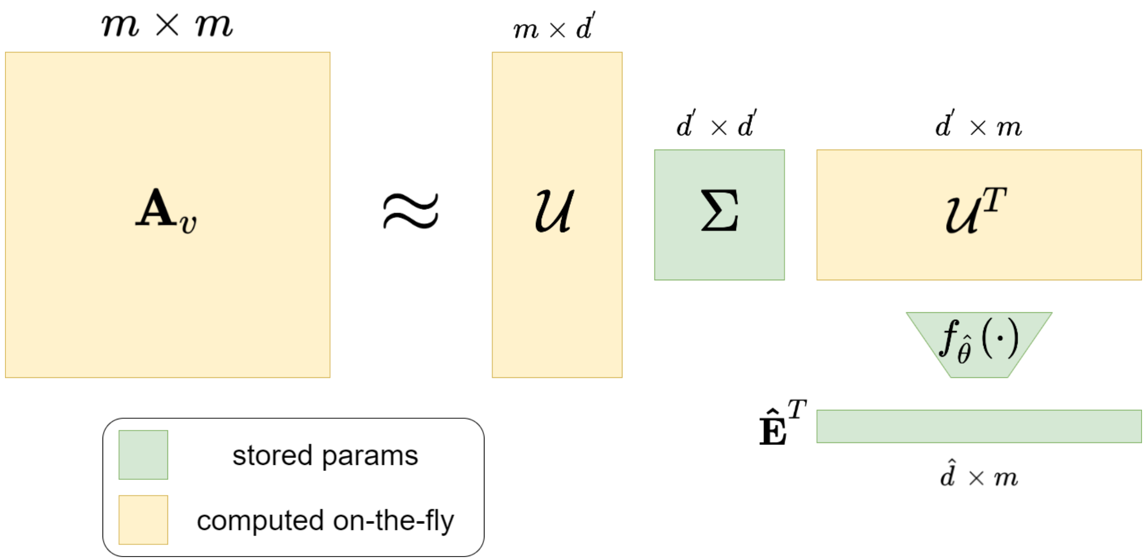
Methodology

Dimension Explosion

For online advertisement systems

#Field $n \leq 100$

#Value $m \approx 10^7$



Selection Tensor Decomposition

$$\boldsymbol{A}^t = \alpha \boldsymbol{A}_f^t + (1 - \alpha)\boldsymbol{A}_v^t$$

$\boldsymbol{A}_f^t$ is the field-level selection
$\boldsymbol{A}_v^t$ is the value-level selection
$\alpha \in \{0, 1\}^{\{C_n^t\}}$ is the hybrid tensor

Hybrid-grained Selection

End-to-end selection

Selection results are discrete

$$\boldsymbol{S}(x) = \mathbf{1}_{x>0}, \qquad \frac{dS}{dx} = 1$$

Sparsification-based Selection Alg.

Result

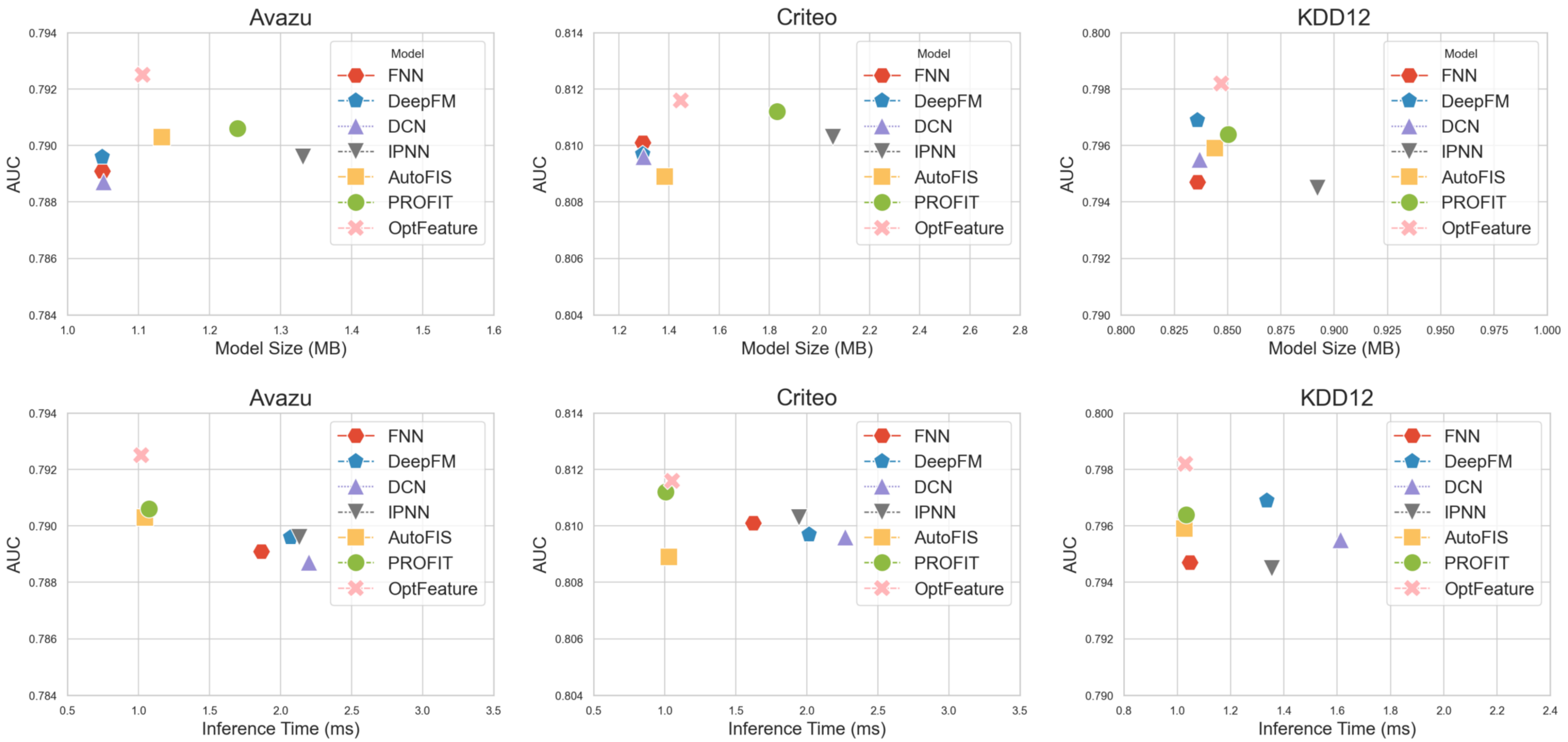
Table 1: Overall Performance Comparison

Dataset		Criteo		Avazu		KDD12	
Category	Model	AUC	Logloss	AUC	Logloss	AUC	Logloss
Shallow	LR	0.7882	0.4609	0.7563	0.3928	0.7411	0.1637
	FM	0.8047	0.4464	0.7839	0.3783	0.7786	0.1566
DSNs	FNN	0.8101	0.4414	0.7891	0.3762	0.7947	0.1536
	DeepFM	0.8097	0.4418	0.7896	0.3757	<u>0.7969</u>	<u>0.1531</u>
	DCN	0.8096	0.4419	0.7887	0.3767	0.7955	0.1534
	IPNN	0.8103	0.4413	0.7896	0.3741	0.7945	0.1537
DSNs with FIS	AutoFIS	0.8089	0.4428	0.7903	0.3749	0.7959	0.1533
	PROFIT	<u>0.8112</u>	<u>0.4406</u>	<u>0.7906</u>	<u>0.3756</u>	0.7964	0.1533
	OptFeature	**0.8116**	**0.4402**	**0.7925***	**0.3741***	**0.7982***	**0.1529***

Here * denotes statistically significant improvement (measured by a two-sided t-test with p-value < 0.05) over the best baseline. The best and second best performed results are marked in **bold** and <u>underline</u> format

On All datasets, OptFeature ranks the first
On Avazu and KDD12, OptFeature achieve significant improvement

Result



OptFeature is both efficient and effective

Summary and Limitations

Summary:

1. Granularity: field -> value -> hybrid.
2. OptFeature:
    1. Selection Tensor Decomposition
    2. Hybrid-grained Selection
    3. Sparsification-based Selection Algorithm
3. Superior in efficiency and effectiveness.

Summary and Limitations

Summary:
1. Granularity: field -> value -> hybrid.
2. OptFeature:
   1. Selection Tensor Decomposition
   2. Hybrid-grained Selection
   3. Sparsification-based Selection Algorithm
3. Superior in efficiency and effectiveness.

Limitations:
1. Lack online evaluation
2. Feature Selection is excluded
3. Single metric-driven

Thanks for Listening!

Summary:
1. Granularity: field -> value -> hybrid.
2. OptFeature:
    1. Selection Tensor Decomposition
    2. Hybrid-grained Selection
    3. Sparsification-based Selection Algorithm
3. Superior in efficiency and effectiveness.

Paper: Code: Myself: