Feature Interactions in Wide and Deep Models for Recommender Systems

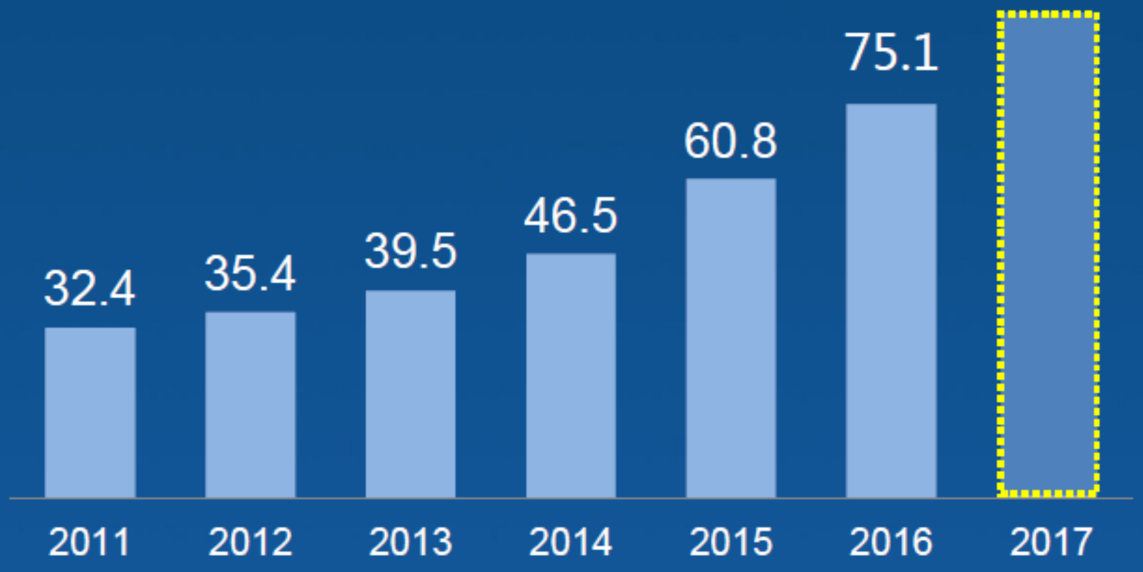
Presenter: Ruiming Tang
Recommendation and Search Lab
Noah's Ark Lab

HUAWEI TECHNOLOGIES CO., LTD.

华为公司概况

180,000 雇员	**80,000** 研发雇员	**170+** 国家	**15** 研发中心	**No. 70** 英特品牌榜单 全球100个品牌上榜	**No. 83** 财富世界500强

华为营业收入（单位: 十亿美元）



*2017年销售收入达922亿美金

运营商业务 ↗24%

- 全球 **NO.1**
- 5G和IoT上的技术先锋

企业业务 ↗ 42%

- 为财富世界500强中的197家企业提供服务

消费者业务 ↗24%

- 品牌知名度, 76% 到 81%
- 出货量: 139 million, ↗ 29%

*过去5年的年平均增长率

华为诺亚方舟实验室专注AI研究

2012实验室	网络智能	企业智能		终端智能	商业成果
诺亚方舟实验室 (350+ patents)	计算视觉	语音语义	推荐搜索	决策推理	先进技术
	AI 理论				
AI研究合作	HIRP	open source	START UP	Professional Advisory Committee	健康生态

10+ 国家，25~大学，50~ 项目，1,000+ 研究人员

全球化布局&本地化研究

华为总部

诺亚实验室全球分布

全球AI能力中心:

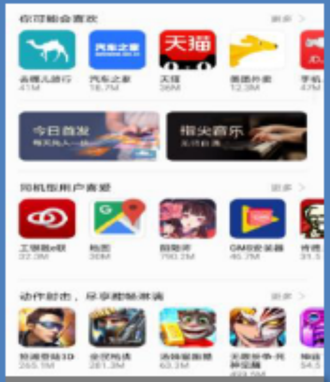
中国: 计算视觉，深度学习，强化学习，决策推理，自然语言处理，AI理论，推荐搜索

北美 & 欧洲: 计算视觉，深度学习，强化学习，决策推理，自然语言处理，AI理论，推荐搜索， 人机交互

推荐搜索

应用

应用商店



游戏中心

新闻推送

华为视频

华为读书

华为音乐

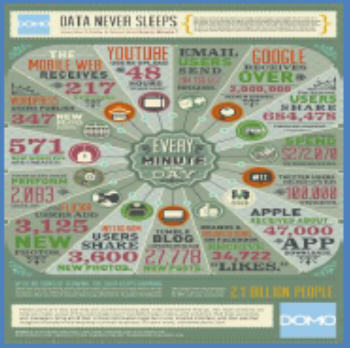
华为云

文档

金融

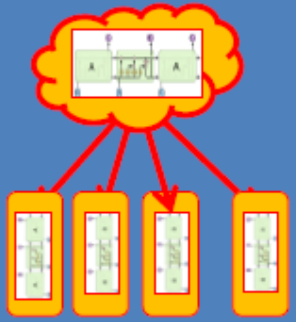
VMall

能力

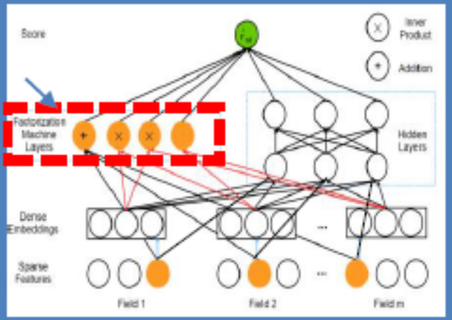


大规模分类

Field-Aware Factorization



端云协同



Deep Factorization Machines

基于联合式元启发学习的对话及推荐系统

Outline

- Recommender Systems and CTR Prediction
- Wide Models for Modeling Feature Interactions
- Deep Models for Modeling Feature Interactions
- AutoML Techniques for Feature Interactions
- Conclusions and Future Directions

Recommender Systems

Music

Video

E-commerce

News

Social network

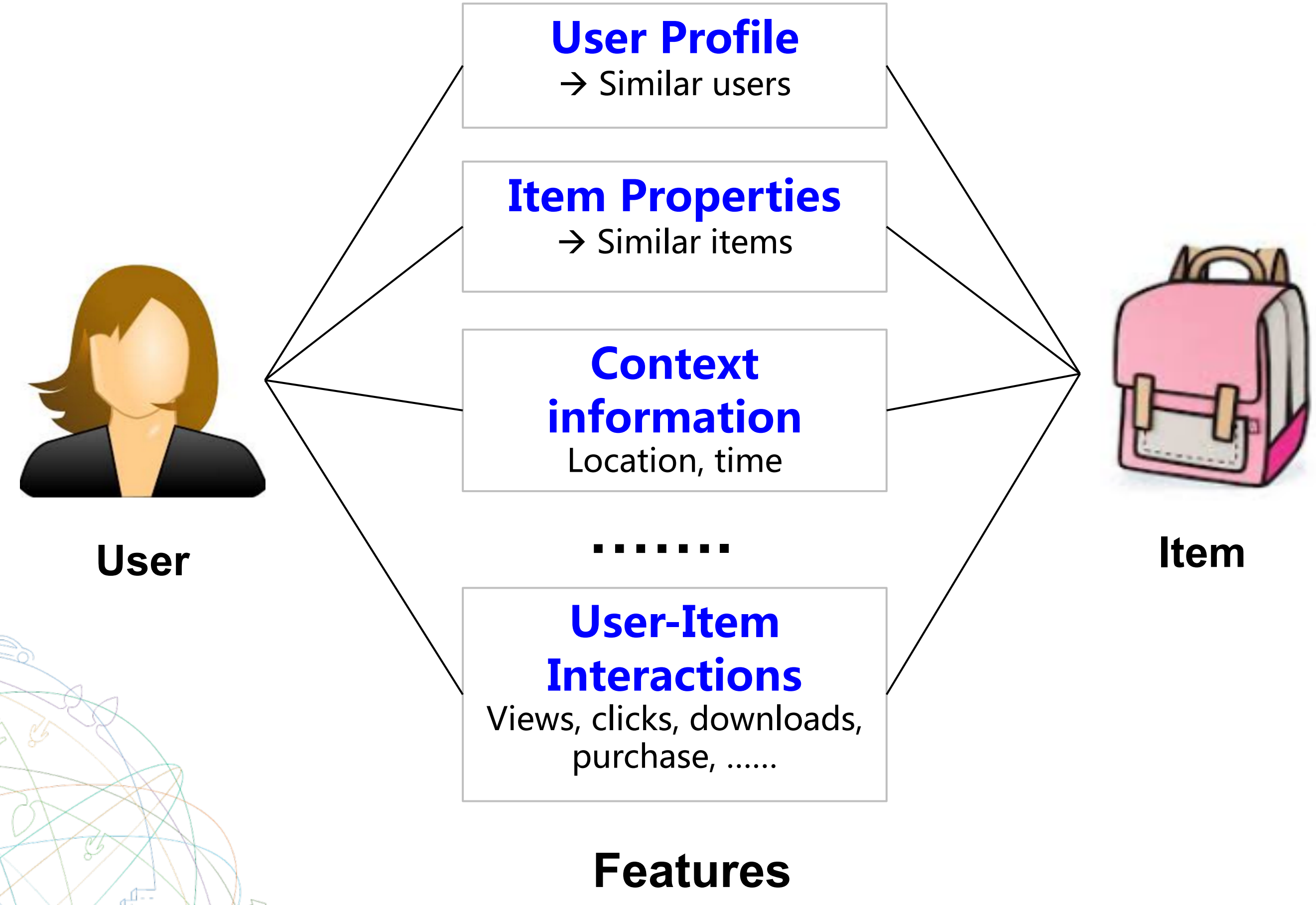
Location-based service

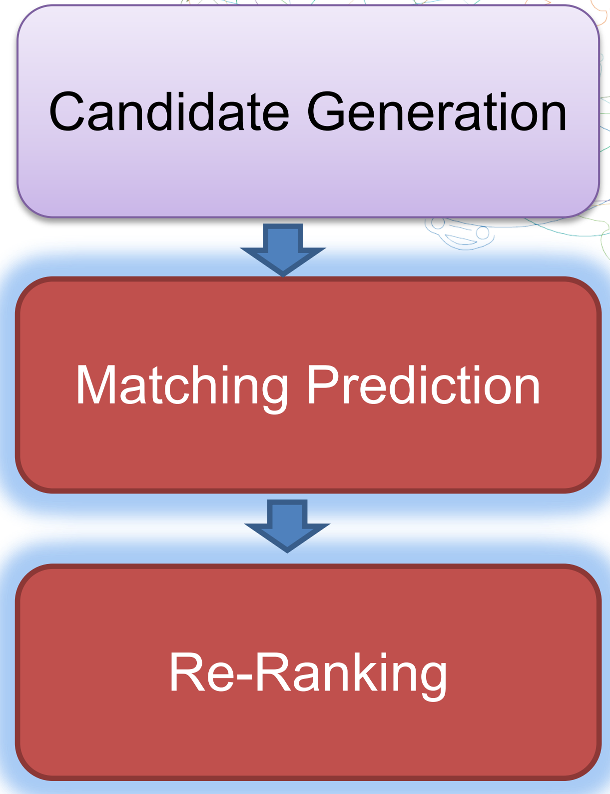
Online Advertising

App

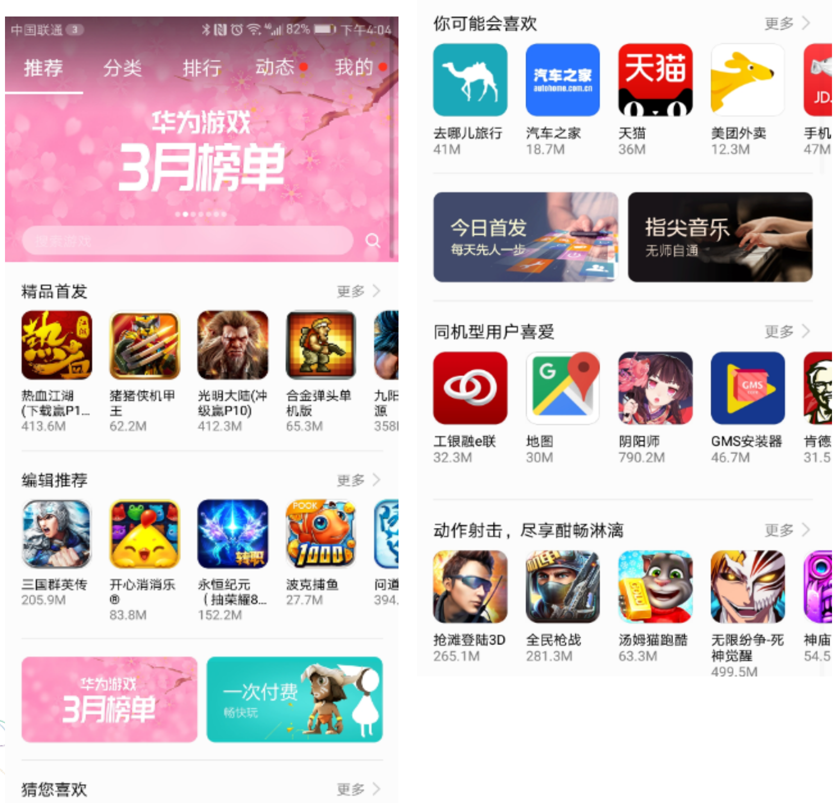
Live broadcasting

Selection + Matching + Re-Ranking



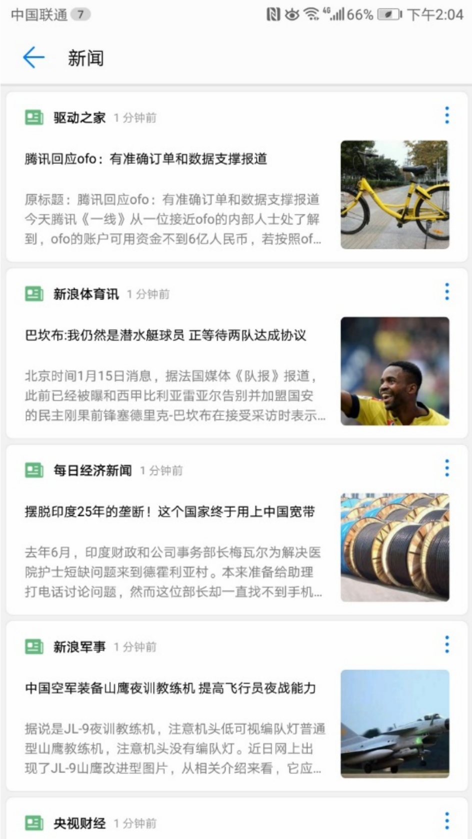


Recommender System in Huawei



App Store and Game Center

HiBoard (service)



News Feed

Road Map of Recommendation Technique

- 2000-2006 Neighborhood based collaborative filtering
- 2007-2009 Matrix factorization and variants
- 2010-2015 Factorization machine and variants
- 2015-2017 Deep neural networks for user behavior prediction
- 2017-2018 Deep reinforcement learning for decision making

CTR Prediction

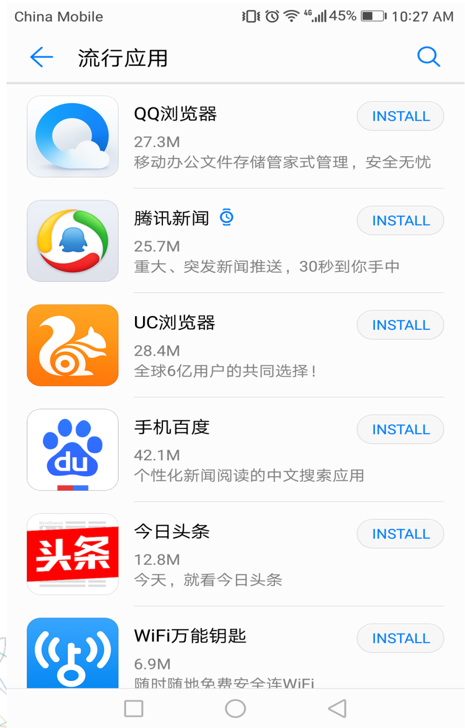
•CTR = Click Through Rate
- ◆ It may be different under different scenarios.

•In online advertising, the advertisements are normally ranked by CTR×Bid:
- ◆CTR represents user satisfaction;
- ◆Bid represents revenue; (different platforms deploy different bidding strategies):
  - ☐Google applies GSP (Generalized Second Price) ;
  - ☐Facebook uses VCG (Vickrey-Clark-Groves) ;

•Other ranking strategies may apply similar idea :
- ◆Game Recommendation: CTR×LTV (Life Time Value) ;
- ◆Video Recommendation: CTR×WT (Watch Time) ;

CTR Prediction in Huawei App Store

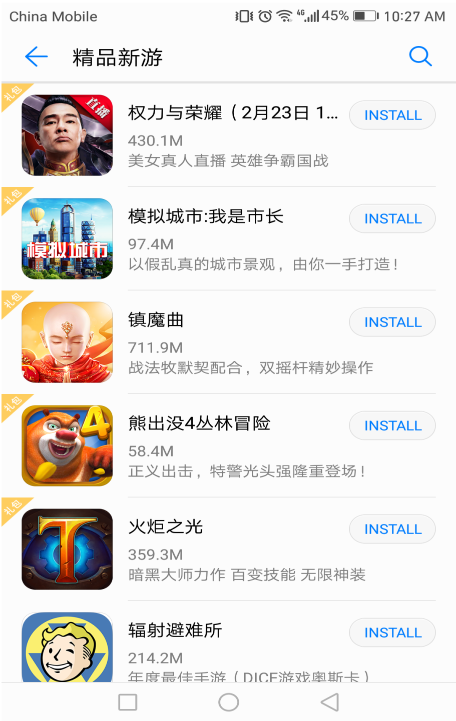
Normal App



Ranked by CTR

CTR -- Click Through Rate

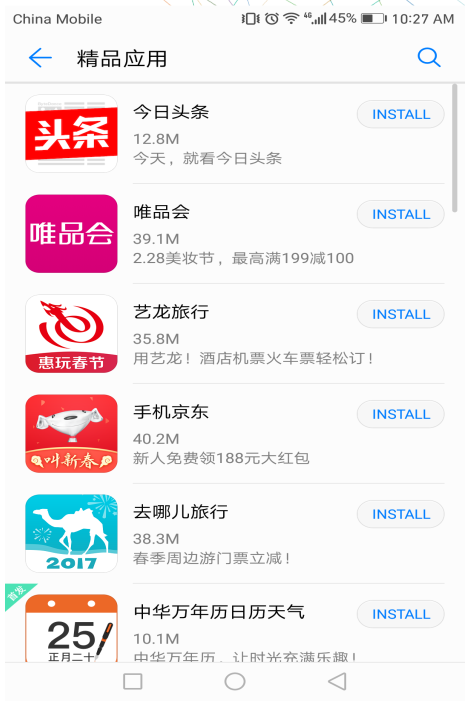
Game App



Ranked by CTR ×LTV

LTV -- Life Time Value

Advertise App



Ranked by CTR ×CPC

CPC – Cost Per Click

Training and Loss Function

In recommendation scenarios, there are categorical features and numerical features. Categorical features are represented by one-hot encoding:

$$x = [\boldsymbol{Weekday = Friday}, \boldsymbol{Gender = Male}, \boldsymbol{City = Shanghai}]$$

$$x = [\mathbf{0, 0, 0, 0, 1, 0, 0} \quad \mathbf{0, 1} \quad \mathbf{0, 0, 1, 0 \ldots 0}]$$

For input instance x_i and a model M with trainable parameters θ, the prediction is $\bar{y_i} = M(x;\theta)$, representing the probability that the input instance x_i leads to a click.

Assume the ground-truth of input x is $y_i \in \{0,1\}$, where $y_i = 1$ means click.

The objective function is to minimize the cross entropy of the predicted values and the ground-truth:

$$\mathcal{L} = \sum_i -y_i \log \bar{y_i} - (1-y) \log(1-\bar{y_i})$$

Feature Interactions Are Important

User click is a complicated behavior to model:

- Both low-order and high-order feature interactions play important roles to model user click behaviors.
  - ✓ People like to download popular apps → id of an app may be a signal for CTR
  - ✓ People often download apps for food delivery at meal time → interaction between app category and time-stamp may be a signal for CTR
  - ✓ Male teenagers like shooting game or RPG → interaction of app category, user gender and age may be a signal for CTR

- Some feature interactions can be easily understood and thus can be designed by experts (like the instances above). Most other feature interactions are hidden in data and difficult to identify (e.g., “diaper and beer” rule). They can be mined automatically by machine learning algorithms.

Outline

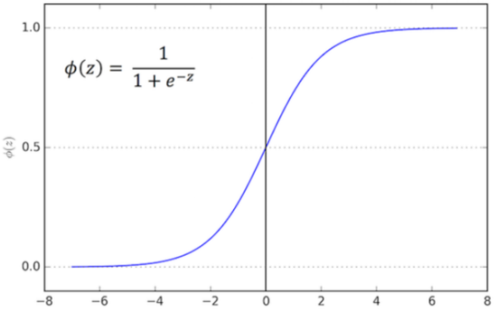
- Recommender Systems and CTR Prediction
- Wide Models for Modeling Feature Interactions
- Deep Models for Modeling Feature Interactions
- AutoML Techniques for Feature Interactions
- Conclusions and Future Directions

Wide Models: LR (Logistic Regression)

- Binary Classification problem.
- The model outputs the click probability.

$$y_{\mathrm{LR}}(x) = sigmoid\left(\sum_{i=1}^{N} w_i x_i\right) = \frac{1}{1+e^{-w^T x}}$$ Prediction function

$$L(y,\hat{y}) = -y\log\hat{y} - (1-y)\log(1-\hat{y}) + \frac{\lambda}{2}\|w\|_2^2$$ Loss function



- Categorical Features are represented by one-hot encoding

$$x = [Weekday = Friday, Gender = Male, City = Shanghai]$$

$$x = [0,0,0,0,1,0,0 \quad 0,1 \quad 0,0,1,0 \ldots 0]$$

- **Feature Interaction**：Non-linear relationship among features are modeled by feature interactions.
  Manually designed.
  Cartesian product → dimension explosion.

  - User.Install ⊗ current Item
  - User.Gender ⊗ Item.Type
  - User.Age ⊗ Item.Type

Tens of millions → Thousands of billions

- The most popular optimization algorithm for LR model is **FTRL**

•Ad Click Prediction: a View from Trenches. In KDD'13, from Google

Wide Models: FM (Factorization Machine)

$$y_{\mathrm{FM}}(\boldsymbol{x}) := \mathrm{sigmoid}\Big(\underbrace{w_0 + \sum_{i=1}^{N} w_i x_i}_{\text{Logistic Regression}} + \underbrace{\sum_{i=1}^{N}\sum_{j=i+1}^{N} \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle x_i x_j}_{\text{Feature Interactions}}\Big)$$

$$y_{\mathrm{FFM}}(\boldsymbol{x}) = \mathrm{sigmoid}\Big(\underbrace{w_0 + \sum_{i=1}^{N} w_i x_i}_{\text{Logistic Regression}} + \underbrace{\sum_{i=1}^{N}\sum_{j=i+1}^{N} \langle \boldsymbol{v}_{i,\mathrm{field}(j)}, \boldsymbol{v}_{j,\mathrm{field(i)}} \rangle x_i x_j}_{\text{Field-aware field embedding}}\Big)$$

•Factorization Machines. In ICDM'10

•Field-aware Factorization Machines for CTR Prediction. In Recsys'16

- Represent each feature by a latent vector;
- The interaction between two features is represented by the inner product of corresponding vectors ;
- *FM is an extend version of MF (Matrix Factorization), with rich side information;*

For x=[Weekday=Friday, Gender=Male, City=Shanghai]

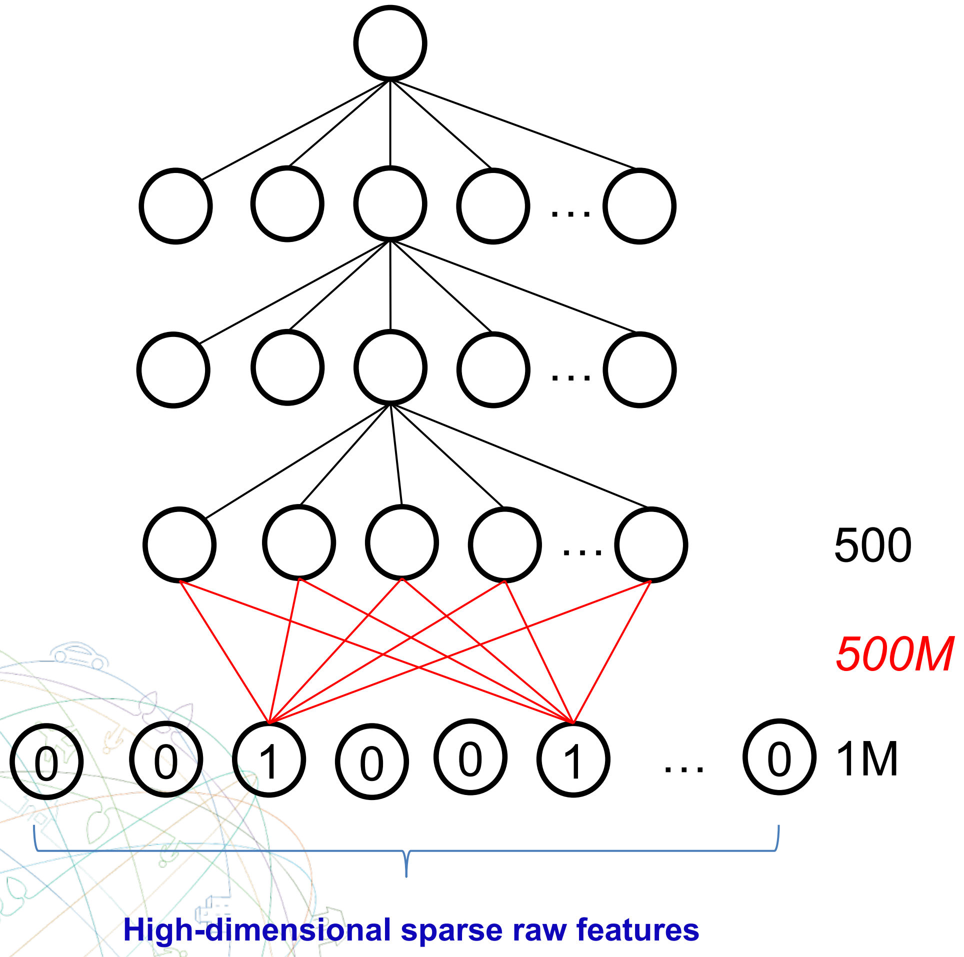
$$y_{\mathrm{MF}}(\boldsymbol{x}) = \mathrm{sigmoid}\Big(w_0 + w_{\mathrm{Friday}} + w_{\mathrm{Male}} + w_{\mathrm{Shanghai}} + \langle \boldsymbol{v}_{\mathrm{Friday}}, \boldsymbol{v}_{\mathrm{Male}} \rangle + \langle \boldsymbol{v}_{\mathrm{Friday}}, \boldsymbol{v}_{\mathrm{Shanghai}} \rangle + \langle \boldsymbol{v}_{\mathrm{Male}}, \boldsymbol{v}_{\mathrm{Shanghai}} \rangle\Big)$$

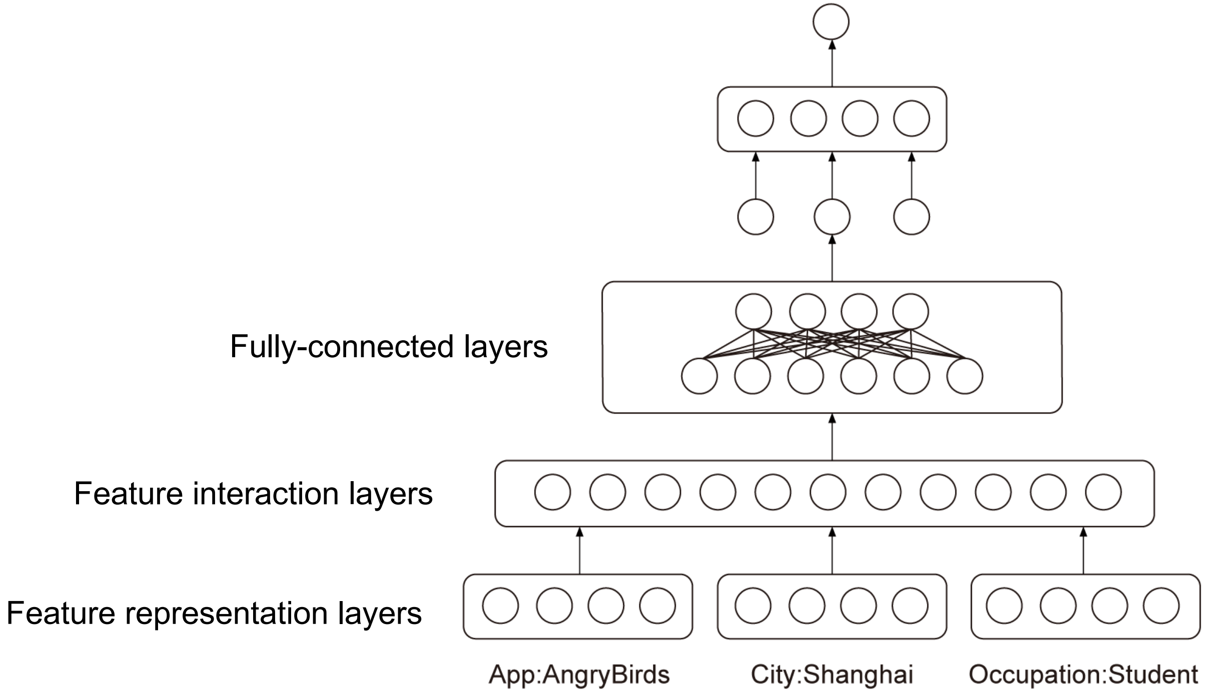
$$y_{\mathrm{FFM}}(\boldsymbol{x}) = \mathrm{sigmoid}\Big(w_0 + w_{\mathrm{Friday}} + w_{\mathrm{Male}} + w_{\mathrm{Shanghai}} + \langle \boldsymbol{v}_{\mathrm{Friday,Gender}}, \boldsymbol{v}_{\mathrm{Male,Weekday}} \rangle + \langle \boldsymbol{v}_{\mathrm{Friday,City}}, \boldsymbol{v}_{\mathrm{Shanghai,Weekday}} \rangle + \langle \boldsymbol{v}_{\mathrm{Male,City}}, \boldsymbol{v}_{\mathrm{Shanghai,Gender}} \rangle\Big)$$

Outline

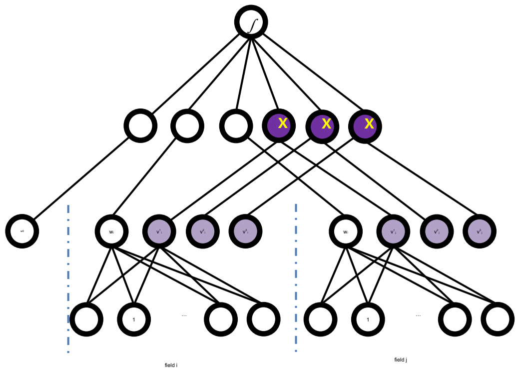
- Recommender Systems and CTR Prediction
- Wide Models for Modeling Feature Interactions
- Deep Models for Modeling Feature Interactions
- AutoML Techniques for Feature Interactions
- Conclusions and Future Directions

Difficulties for Neural Networks in CTR Prediction

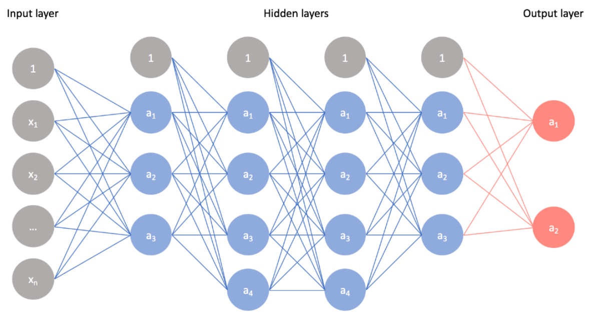




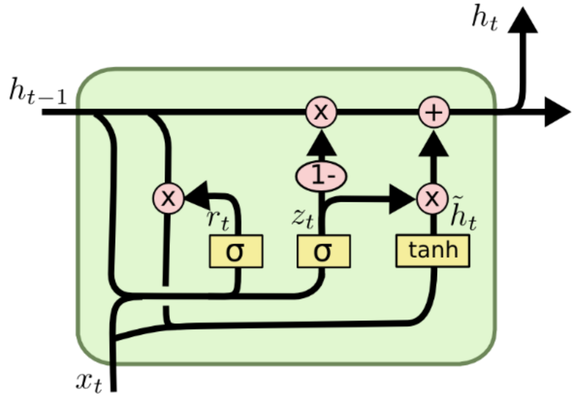
Design of Feature Interactions



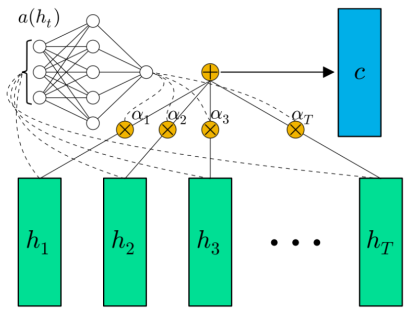
Product operation



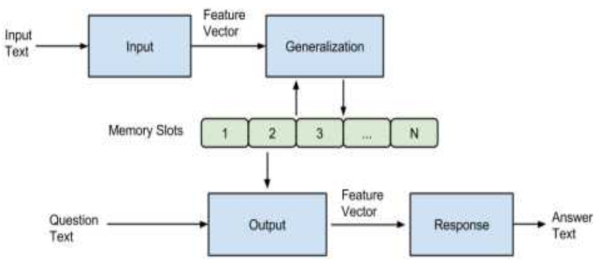
CNN Layer



LSTM\GRU



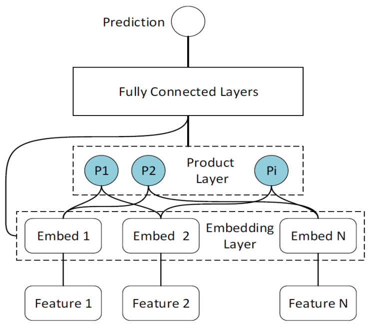
Attention机制



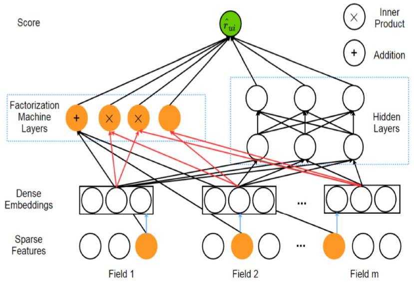
Memory-based network

Product and Attention

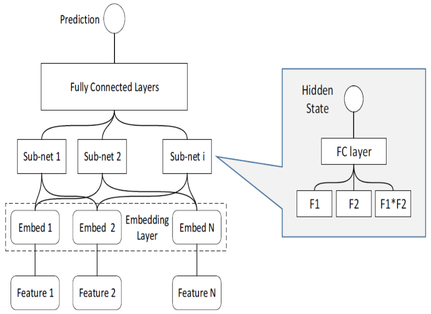
Product Operation



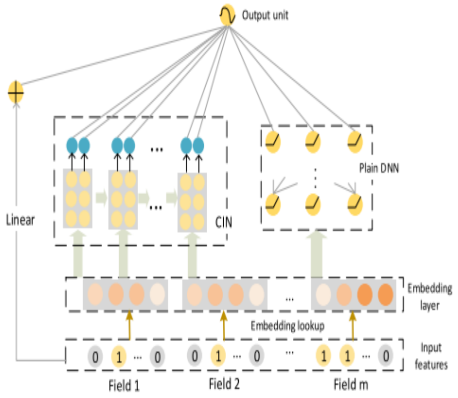
PNN, ICDM2016



DeepFM, IJCAI2017, Huawei

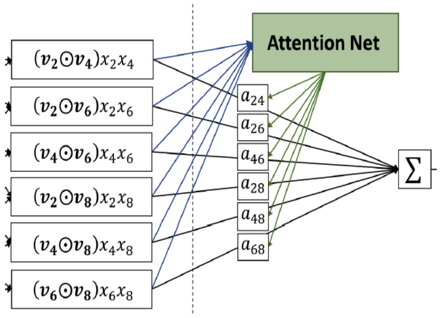


PIN, TOIS2018 , Huawei

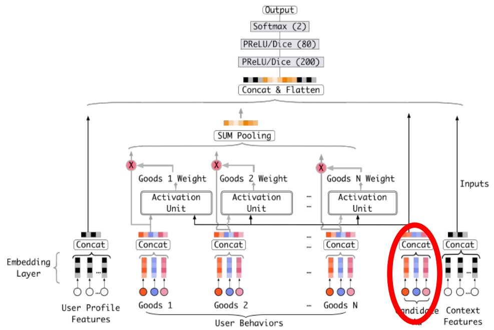


xDeepFM, KDD2018, MSRA

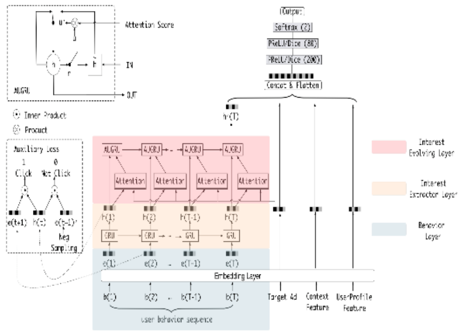
Attention Operation



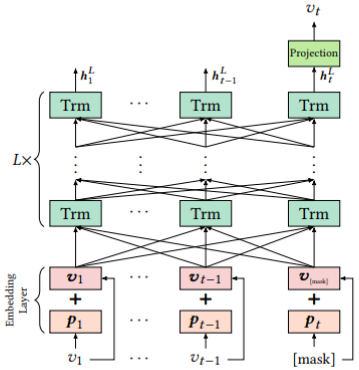
AFM, 2016



DIN, KDD2018 , Alibaba



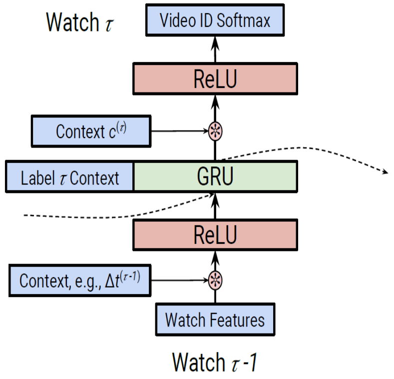
DIEN, AAAI2019 , Alibaba



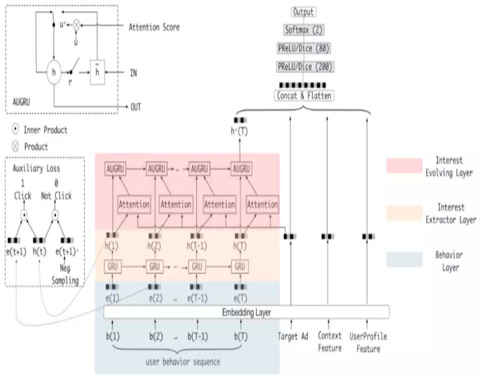
Bert4Rec, 2019 , Alibaba

RNN/CNN and Memory-based

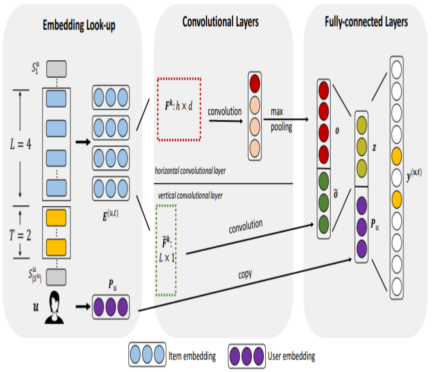
RNN/CNN family



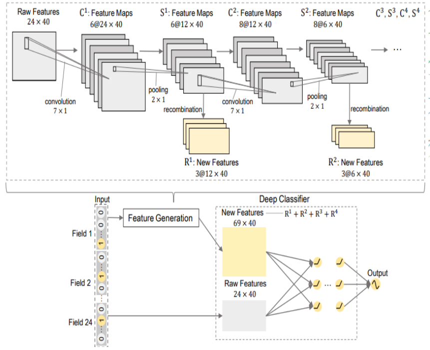
Latent Cross, 2018, Google



DIEN, 2018 , Alibaba

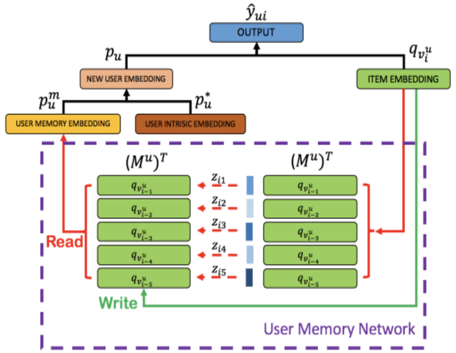


Caser, 2018

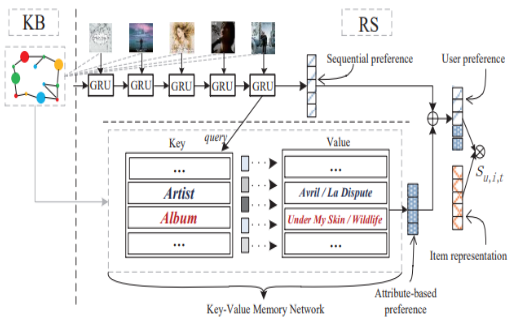


FGCNN , WWW2019 , Huawei

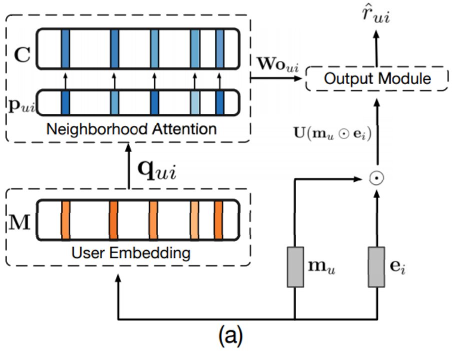
Memory-based



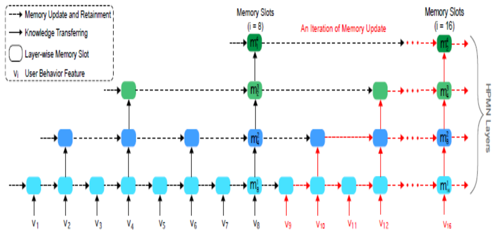
RUM, WSDM2018



KV-MN, SIGIR2018

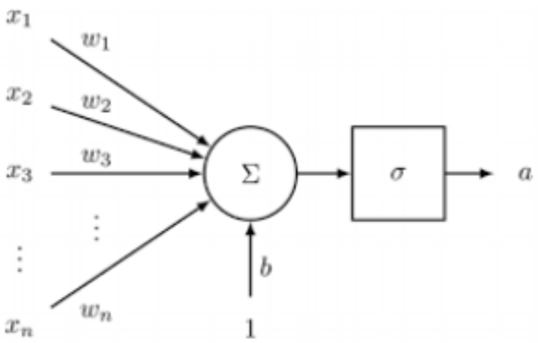


CMN, SIGIR2019

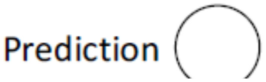


HPMN, 2019 , Alibaba

PNN



Neuron operations in traditional additive networks：
Weighted sum → activation function



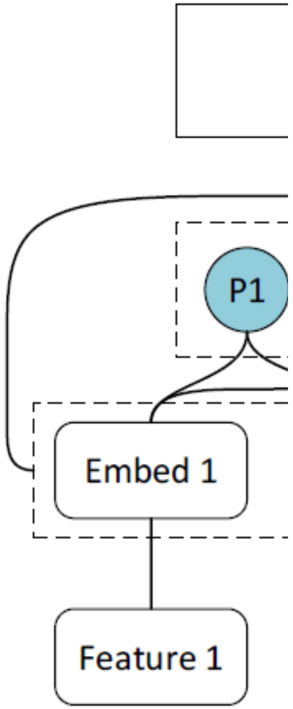
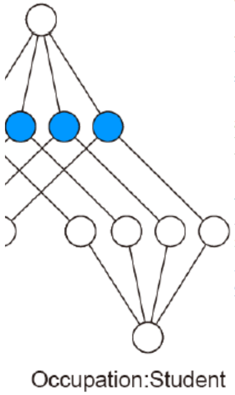


TABLE I: Overall Performance on the Criteo Dataset.

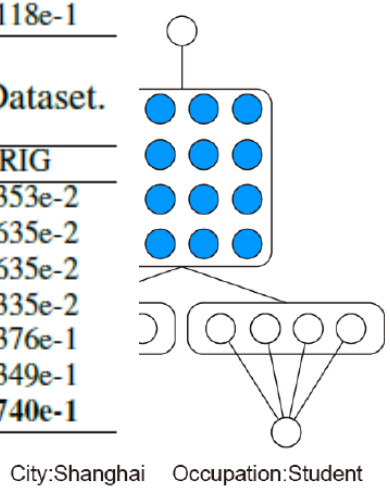
Model	AUC	Log Loss	RMSE	RIG
LR	71.48%	0.1334	9.362e-4	6.680e-2
FM	72.20%	0.1324	9.284e-4	7.436e-2
FNN	75.66%	0.1283	9.030e-4	1.024e-1
CCPM	76.71%	0.1269	8.938e-4	1.124e-1
IPNN	**77.79%**	**0.1252**	**8.803e-4**	**1.243e-1**
OPNN	77.54%	0.1257	8.846e-4	1.211e-1
PNN*	77.00%	0.1270	8.988e-4	1.118e-1

TABLE II: Overall Performance on the iPinYou Dataset.

Model	AUC	Log Loss	RMSE	RIG
LR	73.43%	5.581e-3	5.350e-07	7.353e-2
FM	75.52%	5.504e-3	5.343e-07	8.635e-2
FNN	76.19%	5.443e-3	5.285e-07	9.635e-2
CCPM	76.38%	5.522e-3	5.343e-07	8.335e-2
IPNN	79.14%	5.195e-3	4.851e-07	1.376e-1
OPNN	**81.74%**	5.211e-3	5.293e-07	1.349e-1
PNN*	76.61%	**4.975e-3**	**4.819e-07**	**1.740e-1**



Inner product



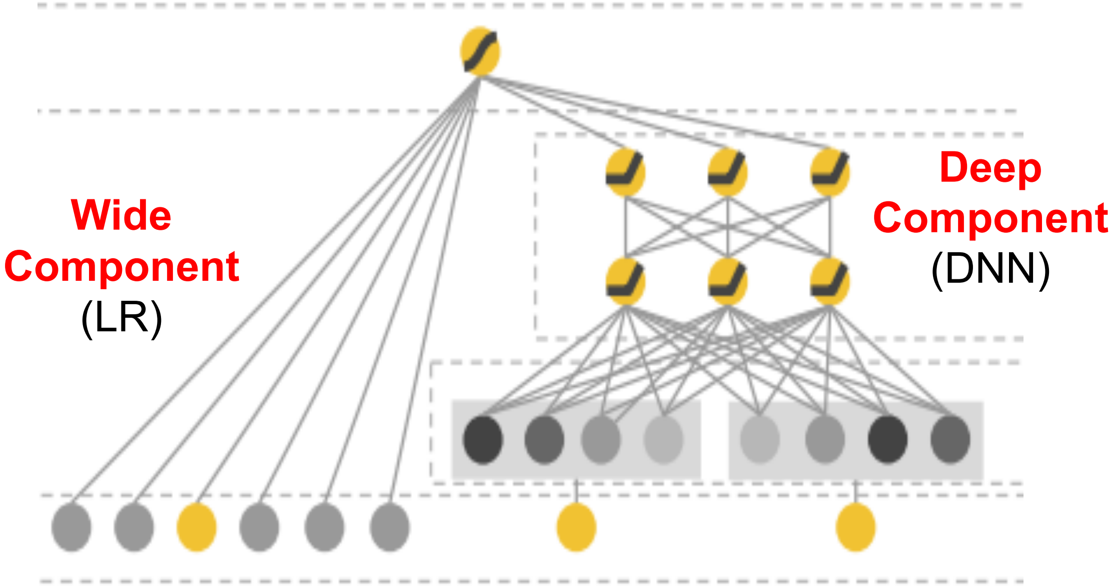
Outer product

✓ **Feature interaction is represented by Product Operation.**

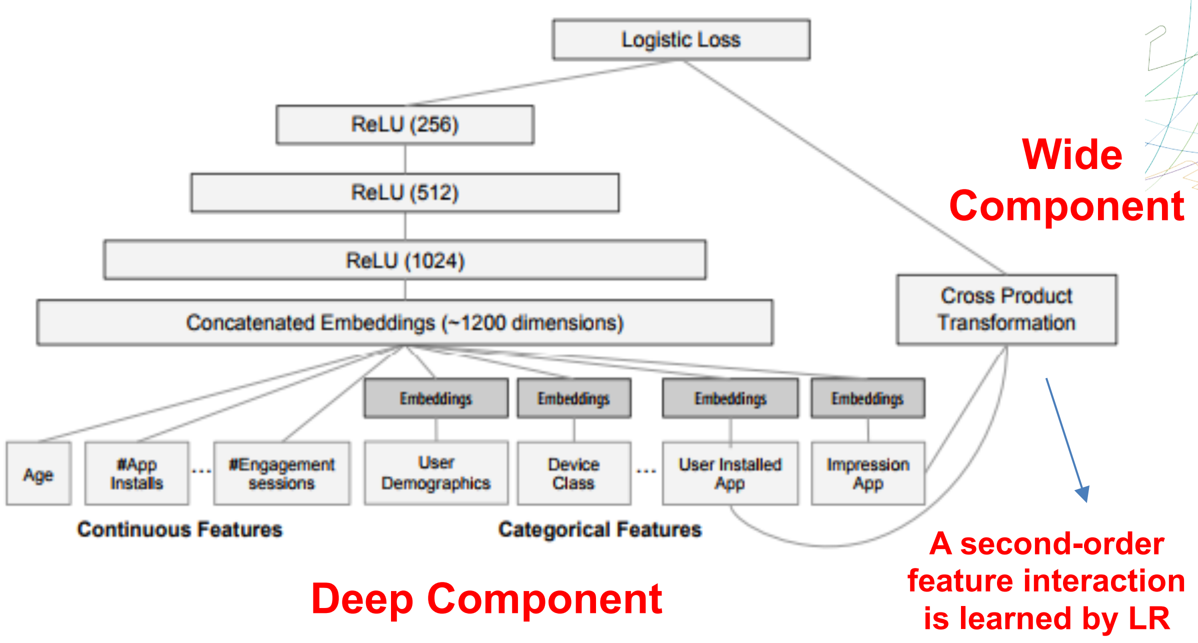
Wide & Deep Learning for Recommendation Systems. In Recsys'16 workshop

Wide & Deep

Wide & Deep Learning

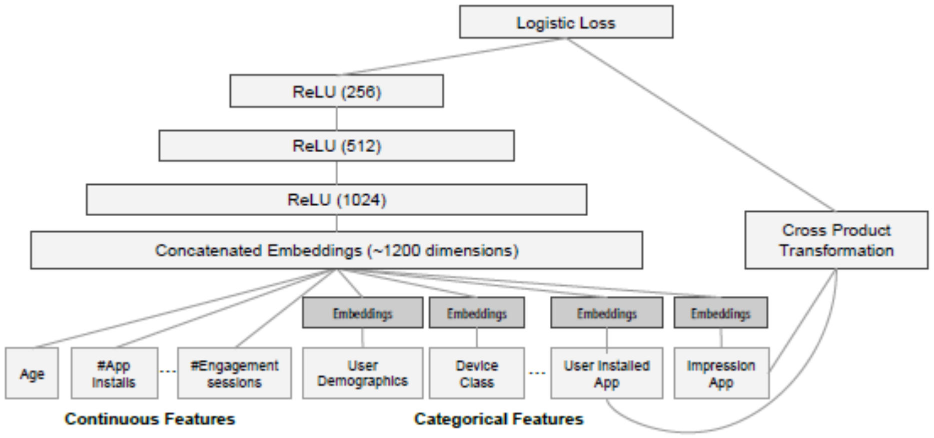


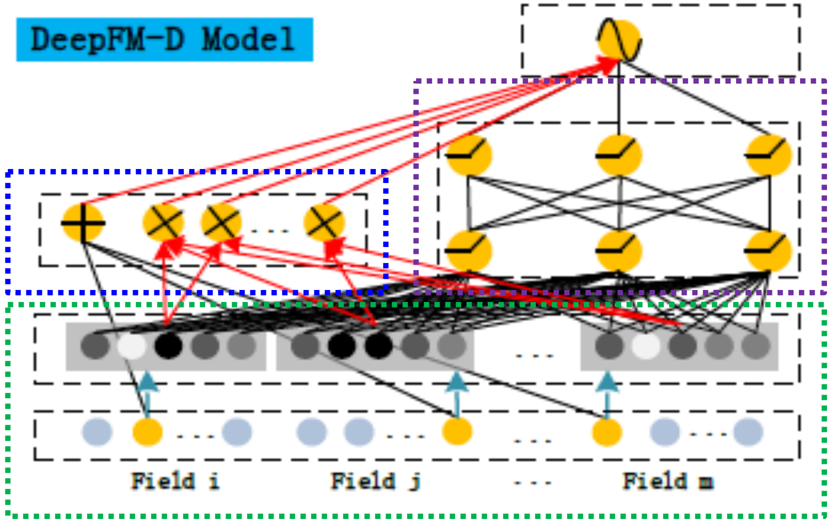
Google Play



Model	Offline AUC	Online Acquisition Gain
Wide (control)	0.726	0%
Deep	0.722	+2.9%
Wide & Deep	0.728	+3.9%

DeepFM





Limitation of Wide & Deep Model:

- Feature engineering is need for the input of Wide Component, because of using LR model.
- What if using FM model as "Wide"? In the experiments, we also try replacing LR with FM (but the input data for them are separated).

DeepFM:

- Using FM as the Wide Component, so feature engineering is not need.
- The same embedding supports both FM Component and Deep Component.
  - The parameters of embedding are trained via back-propagation from both Wide Component and Deep Component.

Table 1: comparison of deep models for CTR prediction

	Pre-training	High-order Feature	Low-order Feature	Feature Engineering
FNN	Yes	Yes	No	No
PNN{1,2,3}	No	Yes	No	No
Wide & Deep	No	Yes	Yes	Yes
DeepFM	**No**	**Yes**	**Yes**	**No**

DeepFM

	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**

wide models

deep models

Wide & deep models

腾讯社交广告 Tencent Social Ads

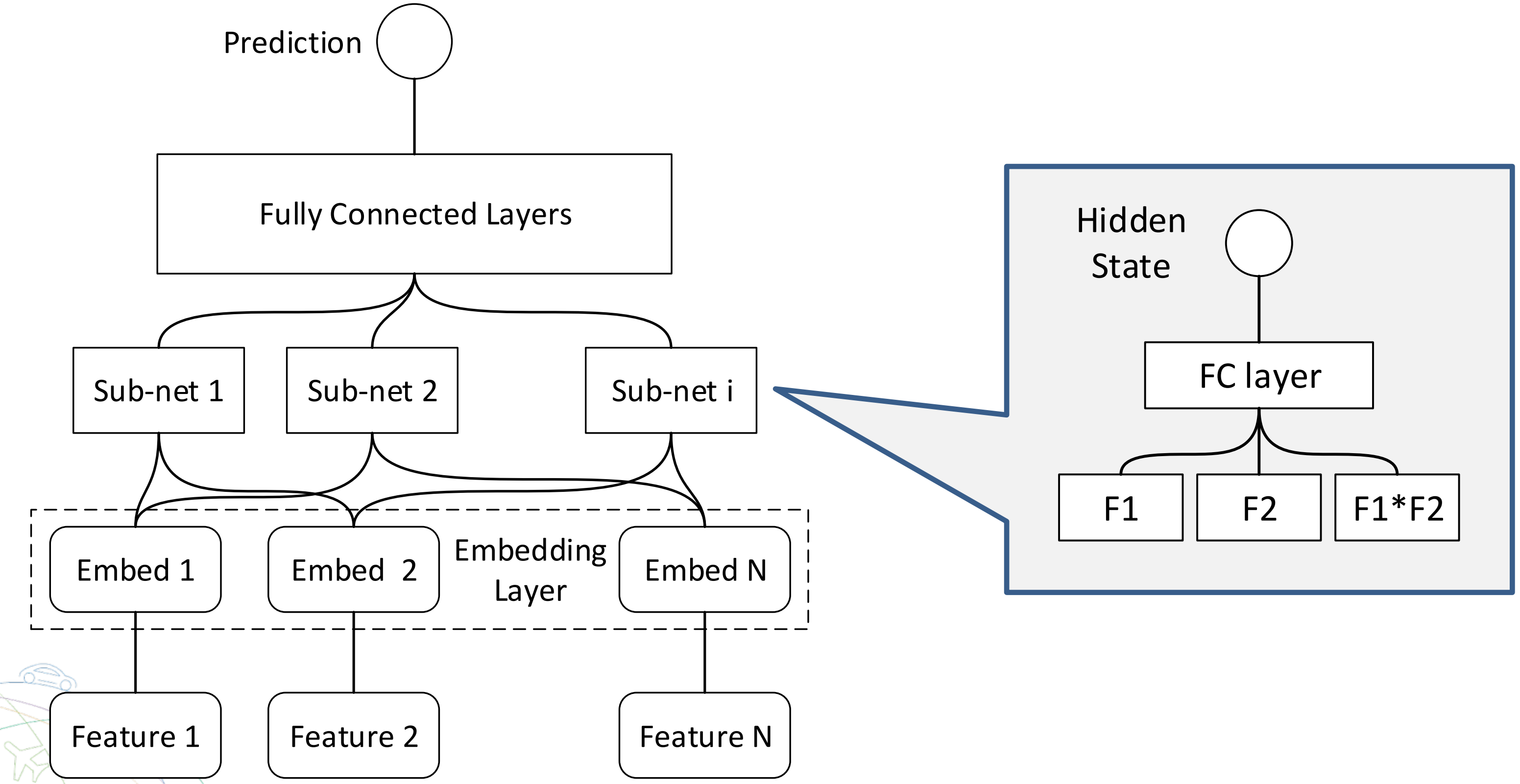
•The Champion team ensembles a set of models, which includes DeepFM model.

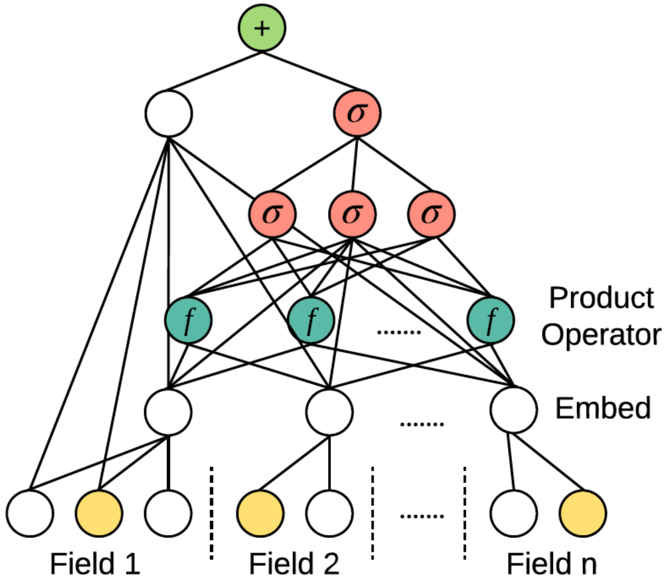
PaddlePaddle

•It opensources our DeepFM.
•Only two deep models for CTR Prediction are implemented by this platform: DeepFM, Wide & Deep.

Product-based Neural Networks for User Response Prediction over Multi-Field Categorical Data(TOIS 2019,Huawei)

PIN





PIN

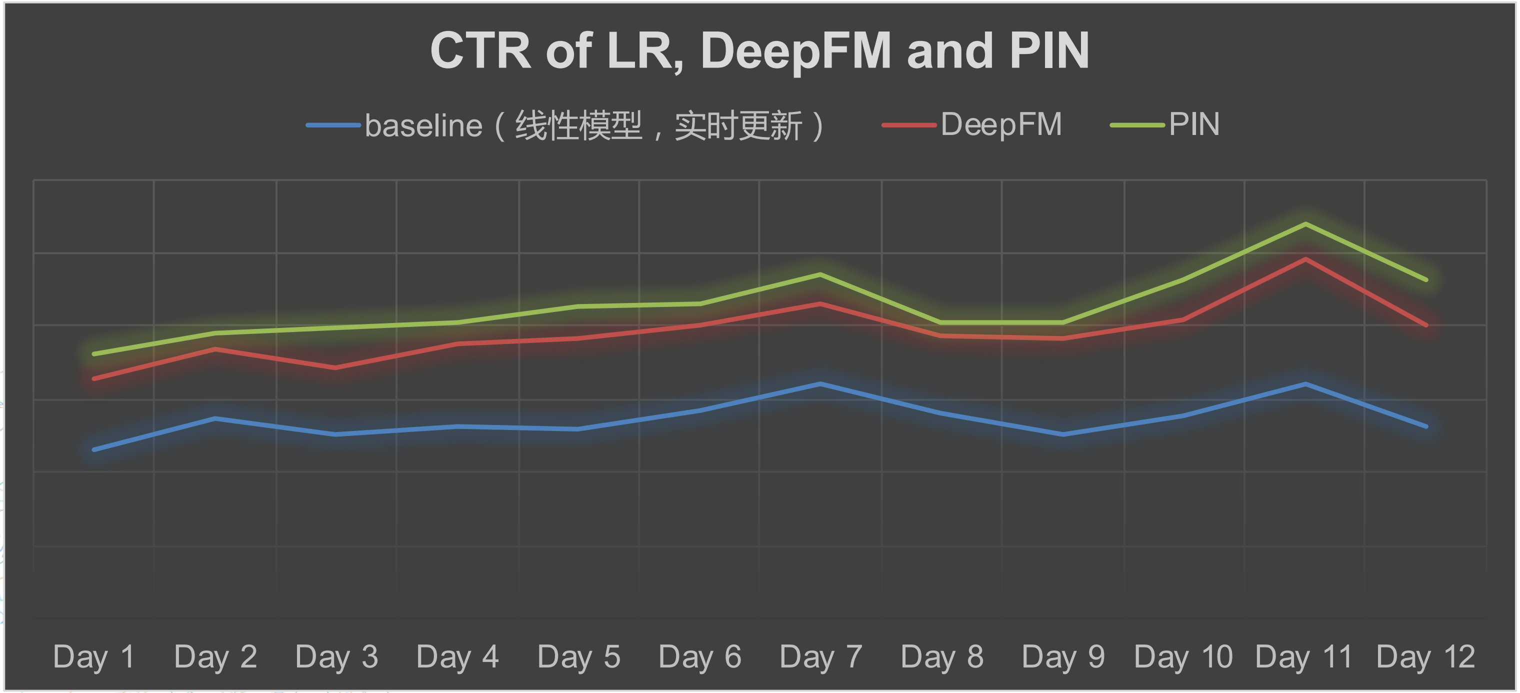
Table 7. Overall performance. (Left-Right: Criteo, Avazu, iPinYou, Huawei)

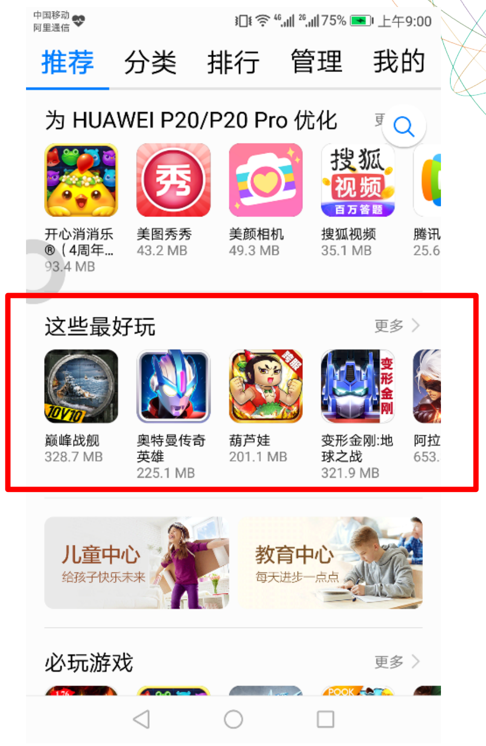
Model	AUC (%)	Log Loss	AUC (%)	Log Loss	AUC (%)	Log Loss	AUC (%)	Log Loss
LR	78.00	0.5631	76.76	0.3868	76.38	0.005691	86.40	0.02648
GBDT	78.62	0.5560	77.53	0.3824	76.90	0.005578	86.45	0.02656
FM	79.09	0.5500	77.93	0.3805	77.17	0.005595	86.78	0.02633
FFM	79.80	0.5438	78.31	0.3781	76.18	0.005695	87.04	0.02626
CCPM	79.55	0.5469	78.12	0.3800	77.53	0.005640	86.92	0.02633
FNN	79.87	0.5428	78.30	0.3778	77.82	0.005573	86.83	0.02629
AFM	79.13	0.5517	78.06	0.3794	77.71	0.005562	86.89	0.02649
DeepFM	79.91	0.5423	78.36	0.3777	77.92	0.005588	87.15	0.02618
KFM	79.85	0.5427	78.40	0.3775	76.90	0.005630	87.00	0.02624
NIFM	79.80	0.5437	78.13	0.3788	77.07	0.005607	87.16	0.02620
IPNN	80.13	0.5399	78.68	0.3757	78.17	0.005549	87.27	0.02617
KPNN	80.17	0.5394	78.71	0.3756	78.21	0.005563	87.28	0.02617
PIN	**80.21**	**0.5390**	**78.72**	**0.3755**	**78.22**	**0.005547**	**87.30**	**0.02614**

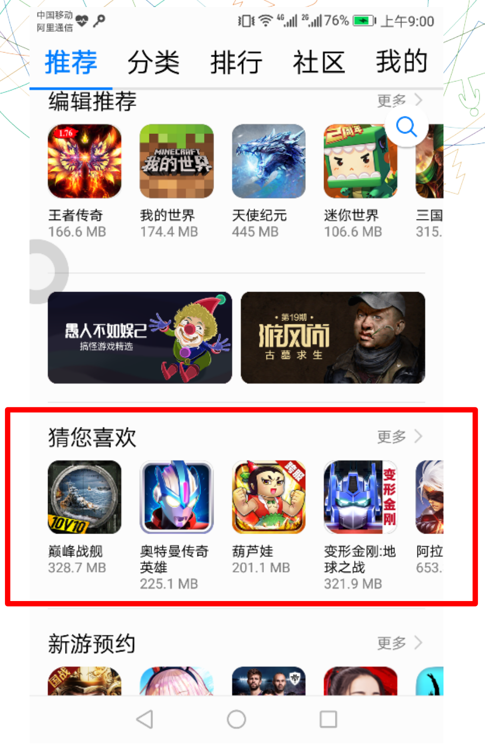
Accepted by TOIS 2019

DeepFM and PIN: Online Experiments

Scenario: “fun games” in Huawei App Market and “guess you like” in Huawei Game Center
Test users: 1% → 10% →30% → 90%
Metric: CTR
Conclusion: Improve CTR by more than 30% in average.



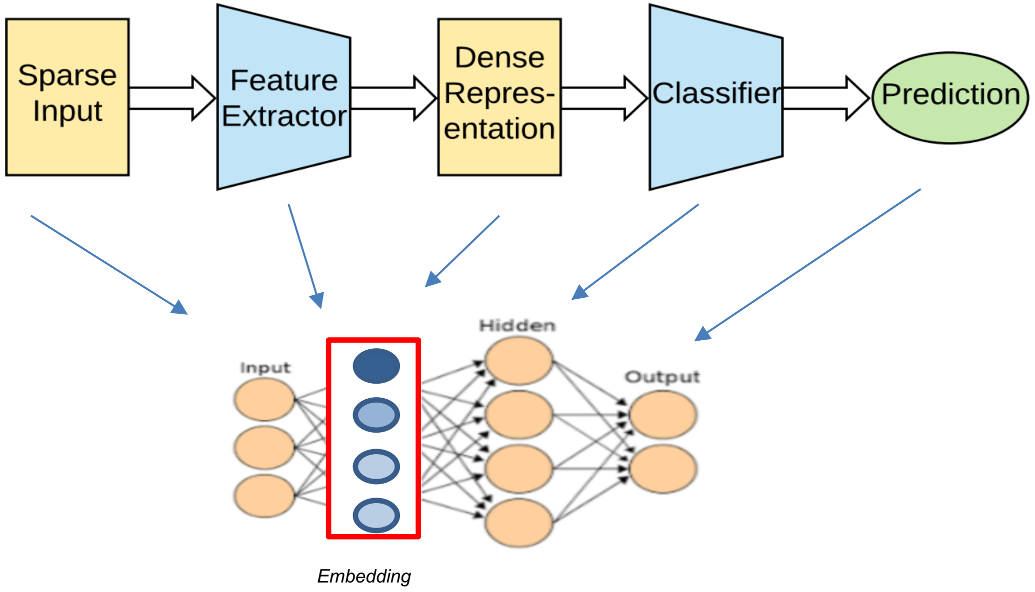




Outline

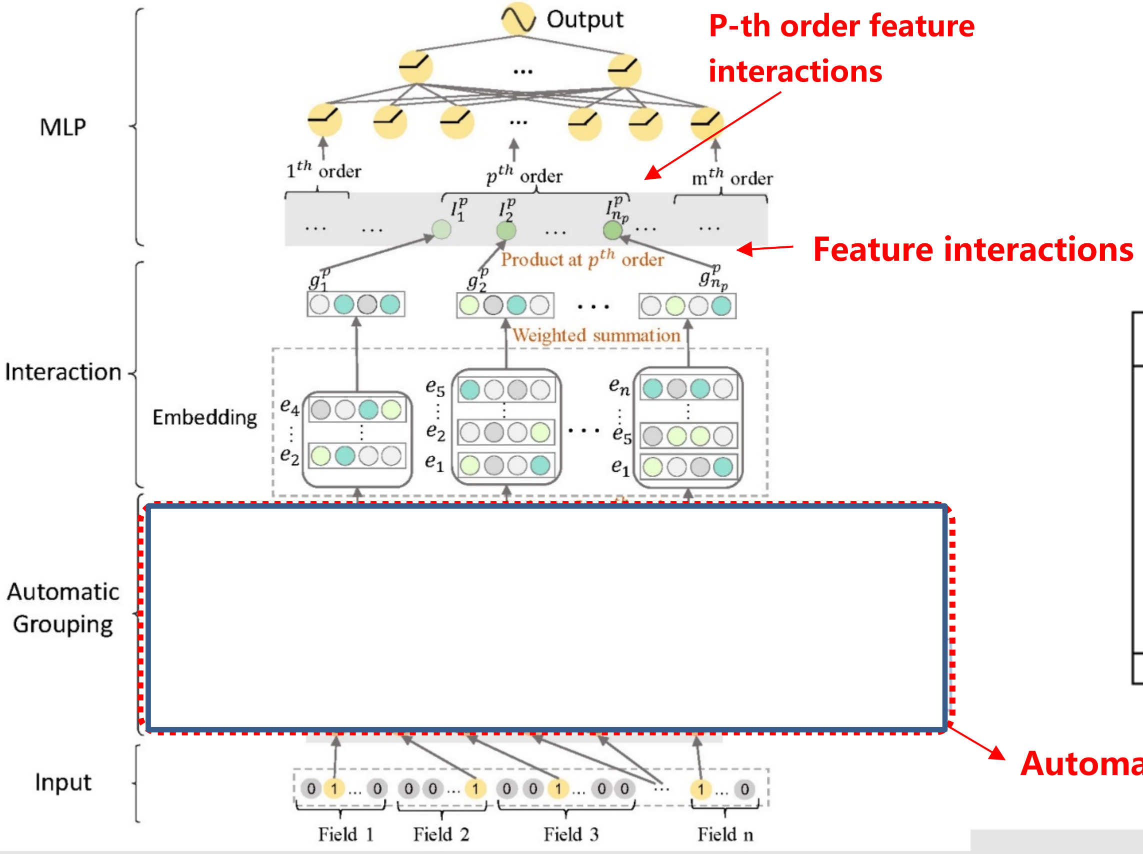
- Recommender Systems and CTR Prediction
- Wide Models for Modeling Feature Interactions
- Deep Models for Modeling Feature Interactions
- AutoML Techniques for Feature Interactions
- Conclusions and Future Directions

AutoML in Deep Learning Models for RecSys



- **Architecture of deep learning models for RecSys: *Embedding + Feature Interaction + MLP***
- **AutoML techniques can be applied in all such three components.**

AutoGroup: Automatic Feature Grouping for Modeling High-order Feature Interactions



Automatically decide feature groups

- Previous models learn feature interactions by enumerating them all and up to second order.
- We learn high-order feature interactions efficiently, avoiding brute-force enumerating.
  - ✓ For each order of feature interactions, we select several groups of features automatically.
  - ✓ For each such group, generate corresponding order of interactions.

Model	Criteo			Avazu			iPinYou		
	AUC	log loss	RI	AUC	log loss	RI	AUC	log loss	RI
LR	78.00%	0.5631	+3.65%	76.76%	0.3868	+3.35%	76.38%	0.005691	+2.88%
GBDT	78.62%	0.5560	+2.64%	77.53%	0.3824	+2.29%	76.90%	0.005578	+1.55%
FM	79.09%	0.5500	+1.81%	77.93%	0.3805	+1.78%	77.17%	0.005595	+1.52%
FFM	79.80%	0.5438	+0.80%	78.31%	0.3781	+1.22%	76.18%	0.005695	+3.05%
AFM	79.13%	0.5517	+1.93%	78.06%	0.3794	+1.55%	77.71%	0.005562	+0.87%
FNN	79.87%	0.5428	+0.66%	78.30%	0.3778	+1.19%	77.82%	0.005573	+0.90%
DeepFM	79.91%	0.5423	+0.59%	78.36%	0.3777	+1.14%	77.92%	0.005588	+0.97%
IPNN	80.13%	0.5399	+0.23%	78.68%	0.3757	+0.67%	78.17%	0.005549	+0.46%
PIN	80.18%	0.5394	+0.16%	78.72%	0.3755	+0.62%	78.22%	0.005547	+0.41%
xDeepFM	80.06%	0.5408	+0.36%	78.55%	0.3766	+0.87%	78.04%	0.005555	+0.60%
FGCNN	80.22%	0.5389	+0.08%	78.82%	0.3747	+0.45%	77.85%	0.005612	+1.22%
AutoGroup	**80.28%★★**	**0.5384★★**	-	**79.15%★★**	**0.3729★★**	-	**78.59%★**	**0.005528★**	-

HUAWEI

AutoFIS: Automatic Feature Interaction Selection in Factorization Models

- Not all the feature interactions are useful. Useless interactions may bring unnecessary noise and complicate the training process.
- We identify useful feature interactions beforehand, in an automatic way and then make the model focus on learning over such useful interactions.

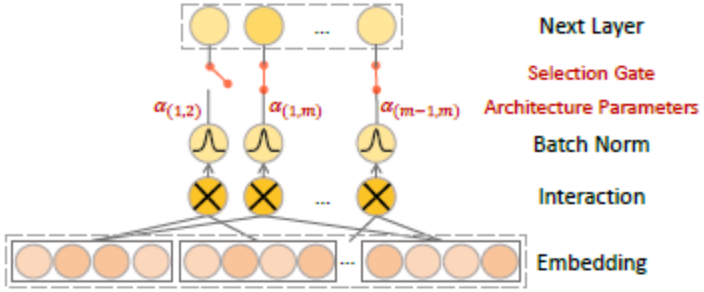
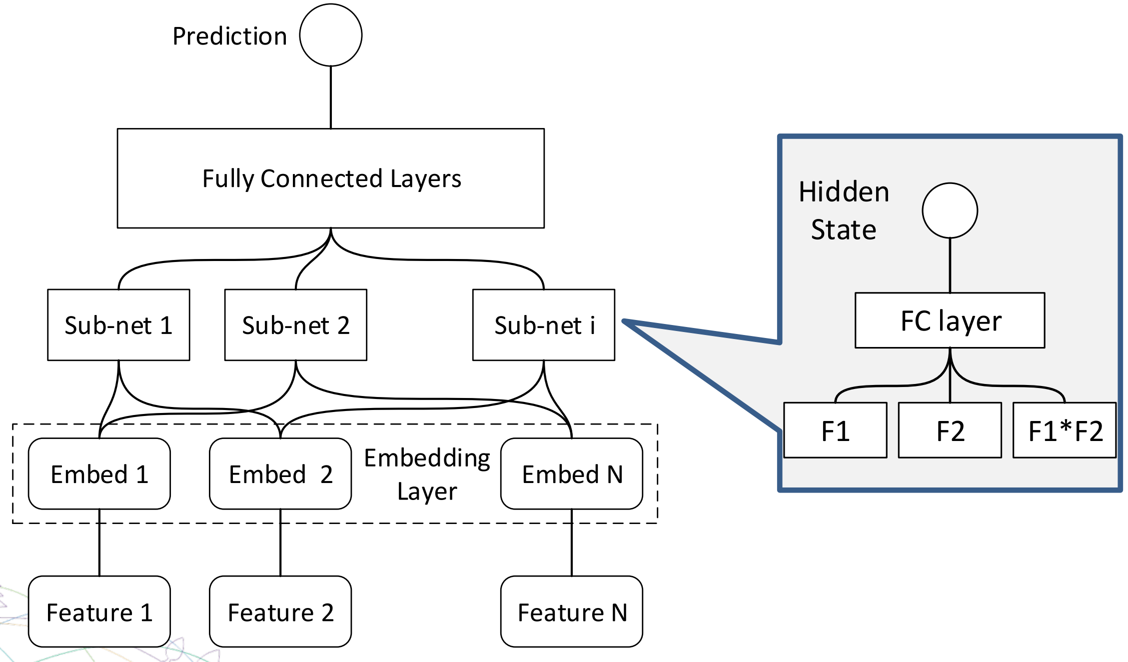
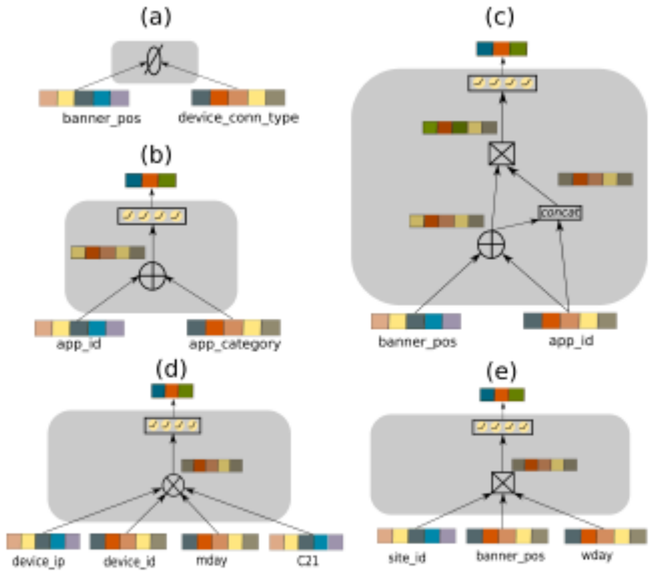


Figure 2: Overview of AutoFIS

Model	Avazu		Criteo		Private	
	AUC	log loss	AUC	log loss	AUC	log loss
FM	0.7793	0.3805	0.7909	0.5500	0.8880	0.08881
FwFM	0.7822	0.3784	0.7948	0.5475	0.8897	0.08826
AFM	0.7806	0.3794	0.7913	0.5517	0.8915	0.08772
FFM	0.7831	0.3781	0.7980	0.5438	0.8921	0.08816
DeepFM	0.7834	0.3776	0.7991	0.5423	0.8948	0.08735
AutoFM	0.7833	0.3777	0.7965	0.5455	0.8944	0.08665
AutoDeepFM	**0.7852**	**0.3766**	**0.8006**	**0.5407**	**0.8979**	**0.08560**
RelaImpr: AutoFM vs FM	0.5%	0.7%	0.7%	0.8%	0.7%	2.4%
RelaImpr: AutoDeepFM vs DeepFM	0.2%	0.3%	0.2%	0.3%	0.3%	2.0%

AutoFeature: Searching for Feature Interactions and Their Architectures





- In PIN model, all the feature interactions are modeled by sub-nets with exactly the same architectures.
- In AutoFeature model, we search automatically different architectures for different subnets.
  - ✓ Complicated feature interactions are modeled by complex architectures.
  - ✓ Simple feature interactions are modeled by trivial architectures.
  - ✓ Useless feature interactions are not modeled.

Outline

- Recommender Systems and CTR Prediction
- Wide Models for Modeling Feature Interactions
- Deep Models for Modeling Feature Interactions
- AutoML Techniques for Feature Interactions
- Conclusions and Future Directions

Conclusions and Future Directions

- Learning Feature Interactions effectively is one of the key factors in Recommender Systems.
- Reinforcement Learning, AutoML and Graph Neural Networks are promising future directions to explore.

Selected Publications

Deep Learning:

- Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, Xiuqiang He: DeepFM: A Factorization-Machine based Neural Network for CTR Prediction. IJCAI 2017
- Huifeng Guo, Ruiming Tang, Yunming Ye, Xiuqiang He: Holistic Neural Network for CTR Prediction. WWW 2017
- Weiwen Liu, Ruiming Tang, Jiajin Li, Jinkai Yu, Huifeng Guo, Xiuqiang He, Shengyu Zhang: Field-aware Probabilistic Embedding Neural Network for CTR Prediction. RecSys 2018
- Bin Liu, Ruiming Tang, Yingzhi Chen, Jinkai Yu, Huifeng Guo, Yuzhou Zhang: Feature Generation by Convolutional Neural Network for Click-Through Rate Prediction. WWW 2019
- Wei Guo, Ruiming Tang, Huifeng Guo, Jianhua Han, Wen Yang, Yuzhou Zhang: Order-aware Embedding Neural Network for CTR Prediction. SIGIR 2019
- Huifeng Guo, Jinkai Yu, Qing Liu, Ruiming, Yuzhou Zhang: PAL: A Position-bias Aware Learning Framework for CTR Prediction in Live Recommender Systems. RecSys 2019
- Yanru Qu, Bohui Fang, Weinan Zhang, Ruiming Tang, Minzhe Niu, Huifeng Guo, Yong Yu, Xiuqiang He: Product-based Neural Network for User Response Prediction over Multi-field Categorical Data. TOIS 2019

Reinforcement Learning:

- Feng Liu, Ruiming Tang, Xutao Li, Yunming Ye, Huifeng Guo, Xiuqiang He: Novel Approaches to Accelerating the Convergence Rate of Markov Decision Process for Search Result Diversification. DASFAA 2018
- Haokun Chen, Xinyi Dai, Han Cai, Weinan Zhang, Xuejian Wang, Ruiming Tang, Yuzhou Zhang, Yong Yu: Large-scale Interactive Recommendation with Tree-structured Policy Gradient. AAAI 2019
- Feng Liu, Huifeng Guo, Xutao Li, Ruiming Tang, Yunmin Ye, Xiuqiang He: End-to-End Deep Reinforcement Learning based Recommendation with Supervised Learning. WSDM 2020

Graph Neural Networks:

- Jianing Sun, Yingxue Zhang, Chen Ma, Mark Coates, Huifeng Guo, Ruiming Tang, Xiuqiang He: Multi-Graph Convolution Collaborative Filtering. ICDM 2019

AutoML:

- All under reviewing process.

联系我们

关注我们: Noah_ark_lab

联系我们：http://www.noahlab.com.hk/

tangruiming@huawei.com