

Interpretable Click-Through Rate Prediction through Hierarchical Attention

Zeyu Li, Wei Cheng, Yang Chen, Haicheng Chen, Wei Wang

UCLA, NEC Labs America, Google

WSDM 2020

Houston, Texas, USA

Recommender systems



淘宝网
Taobao.com

- 女装 / 男装 / 内衣
- 童装 / 箱包 / 配件
- 童装玩具 / 孕产 / 用品
- 家电 / 数码 / 手机
- 美妆 / 洗护 / 保健品
- 珠宝 / 眼镜 / 手表
- 运动 / 户外 / 乐器
- 游戏 / 动漫 / 影视
- 美食 / 生鲜 / 零食
- 鲜花 / 宠物 / 农资
- 房产 / 装修 / 建材
- 家具 / 家纺 / 家纺
- 汽车 / 二手车 / 用品
- 办公 / DIY / 五金电子

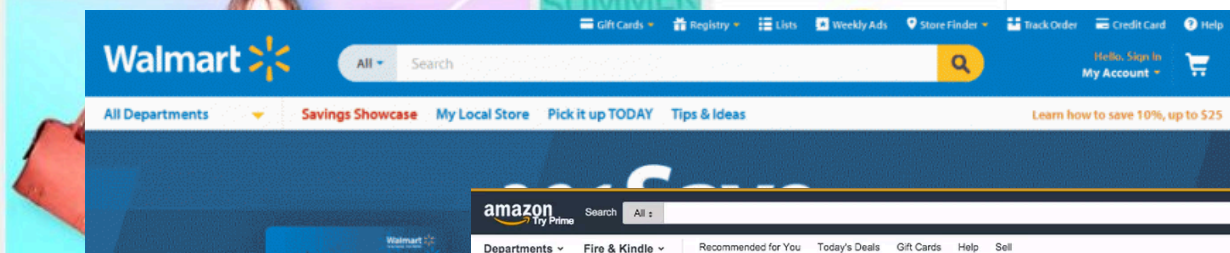


淘宝网 天猫 全球购

全部商品 热门搜索

白瓷牙骨 摄像头 钢琴 生日礼物 篮球 蛋糕 婴儿推车 气球 帽子 爬行垫 水杯 女士凉鞋 电动车

天猫 聚划算 天猫超市 淘抢购 电器城 司法拍卖 中国制造 特色中国 阿里旅行 智能生活 苏宁易购



Walmart

All Departments Savings Showcase My Local Store Pick it up TODAY Tips & Ideas

Learn how to save 10%, up to \$25

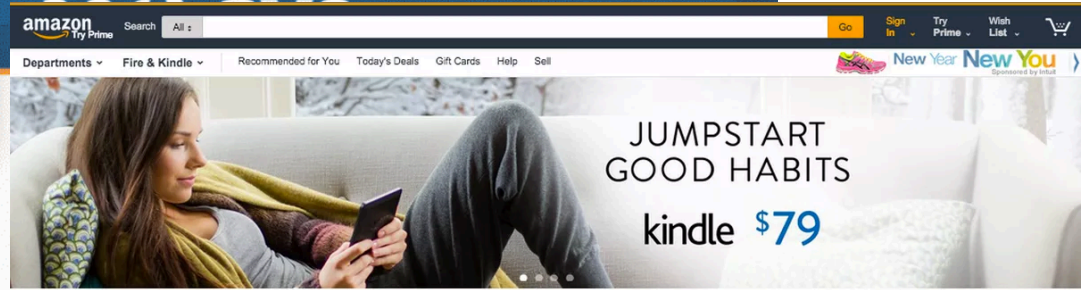


Featured products

Rollback \$299.99 Was \$449.00 Save \$149.01 Dyson Ball Compact Animal Bagless...

Rollback \$669.99 Was \$729.99 Save \$60.00 HP White Pavillon 23-Q120 All-in-One...

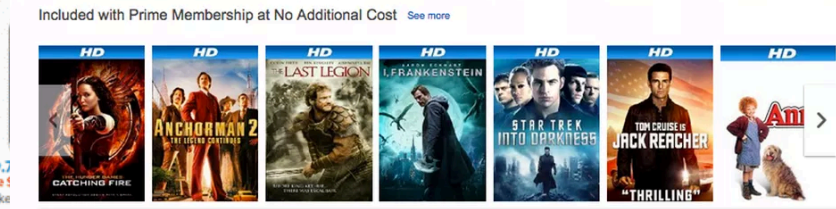
\$69.7 Save! Click! Prepa



amazon Try Prime

Departments Fire & Kindle Recommended for You Today's Deals Gift Cards Help Sell

JUMPSTART GOOD HABITS kindle \$79



Included with Prime Membership at No Additional Cost

CATCHING FIRE ANCHORMAN 2 THE LAST LEGION FRANKENSTEIN STAR TREK INTO DARKNESS JACK REACHER



Commute in Comfort

GARMIN



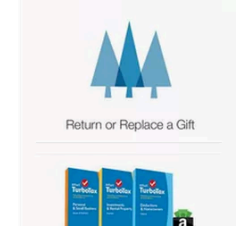
LG Watch R

DESIGN COMES FULL CIRCLE

LG Electronics G Watch R W110 Smart Watch

★★★★★ (29)

\$309.00



Return or Replace a Gift

Click-throughs

Walmart logo and navigation bar. Search bar with 'All' dropdown and 'Search' button. Account links: Hello, Sign In, My Account. Navigation: All Departments, Savings Showcase, My Local Store, Pick it up TODAY, Tips & Ideas. Learn how to save 10%, up to \$25.

3-2-1 Save
A new way to save more every day.
Savings paid as a statement credit.

Featured products

\$299.99 Was \$449.00 Save \$149.01	\$669.99 Was \$729.99 Save \$60.00	\$69.74 Was \$99.00 Save \$30.14	\$10.51	\$14.92	\$299.00	\$139.00 Was \$149.00 Save \$10.00



BLACK FRIDAY DEALS WEEK

Deals on Amazon devices

Echo Spot Black \$89.99 \$129.99	Fire TV Stick 4K w/Remote \$34.99 \$49.99	Fire 7-inch Tablet w/Alexa \$29.99 \$49.99	All New Echo Dot (3rd Gen) \$24.00 \$49.99
Fire 7-inch Kids Tablet \$69.99 \$99.99	All New Echo Plus (2nd Gen) \$109.99 \$149.99	All New Echo Show (2nd Gen) \$179.99 \$229.99	Fire TV Cube w/Alexa \$59.99 \$119.99

淘宝网 Taobao.com

搜索 高级搜索

白灼游骨 馒头头 焖饼 生日礼物 蓝祥 相框 婴儿推车 气球 帽子 骑行垫 水杯 女士凉鞋 电动车

天猫 聚划算 天猫超市 淘抢购 电器城 司法拍卖 中国质造 特色中国 阿里旅行 智能生活 苏宁易购

主题市场

- 女装 / 男装 / 内衣
- 鞋靴 / 箱包 / 配件
- 童装玩具 / 孕产 / 用品
- 家电 / 数码 / 手机
- 美妆 / 洗护 / 保健品
- 珠宝 / 眼镜 / 手表
- 运动 / 户外 / 乐器
- 游戏 / 动漫 / 影视
- 美食 / 生鲜 / 零食
- 鲜花 / 宠物 / 农资
- 房产 / 装修 / 建材
- 家具 / 家纺 / 家纺
- 汽车 / 二手车 / 用品
- 办公 / DIY / 五金电子

时尚时间 潮货汇集 不潮还不赶快点我

半身裙的搭配 N种方案推荐

HI 你好 淘宝金币兑换 聚划算会员专享

登录 注册 开店

公告 规则 论坛 安全 公益

阿里巴巴推出跨境体系 打造未来企业跨境通行证

阿里集团启动数据湖 史上最大规模数据湖

4.8 今日热卖

盛夏超睡节 避暑不涨价

凯立德智能 胎压监测助手

新车特卖活动 PLO500万

名店潮流每日上新 5折起

充值 考试 旅行 保险 彩票 电影 点外卖 理财

Click-throughs



Two natural questions to ask:

1. How many advertisements will be clicked?
2. How many clicks will be purchased?

Click-throughs



Two natural questions to ask:

1. How many advertisements will be clicked?
2. How many clicks will be purchased?

Click-Through Rate Prediction (CTR)

- CTR:
 - Important role in recommendation system
 - Revenue of advertisements



Image: <https://www.lyfemarketing.com/blog/average-click-through-rate/>

Background

- CTR: *binary prediction*
- Pre-Deep Learning Model
 - FM: Factorization Machine
 - MF: Matrix Factorization
 - LR: Logistic Regression
- Deep learning based CTR model
 - DeepFM = FM module + Deep module
 - xDeepFM = CIN module + Deep module
 - CIN: Compressed Interest Network
 - and more ...

SOTA models with DNN

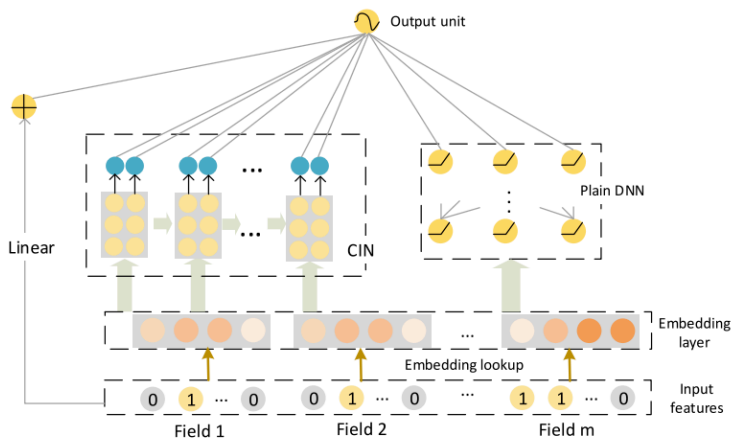


Figure 5: The architecture of xDeepFM.

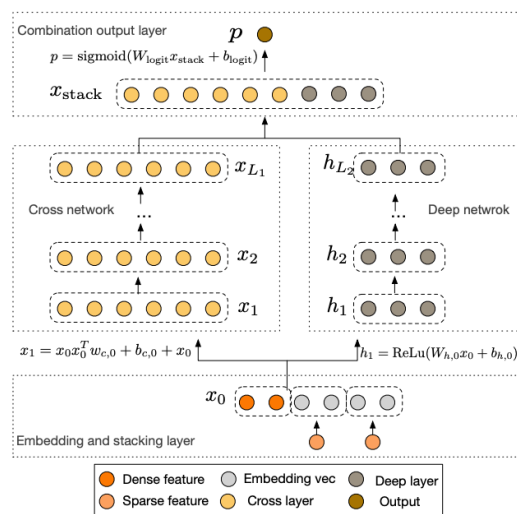
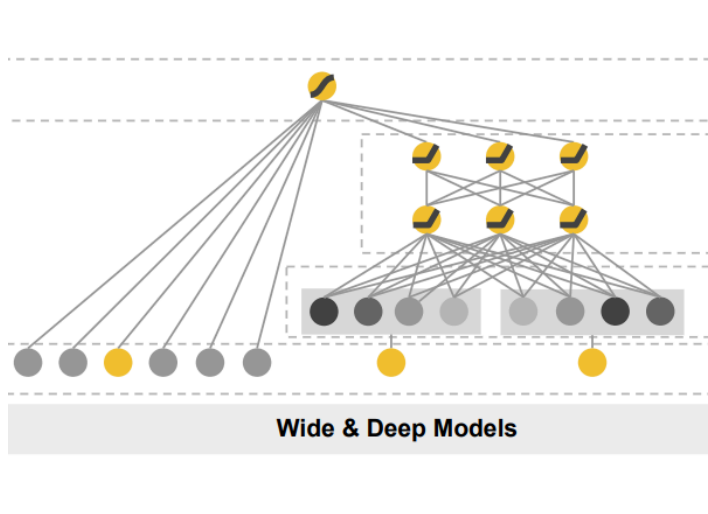
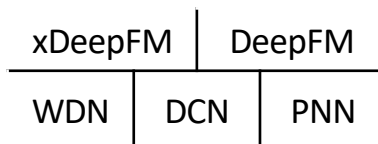


Figure 1: The Deep & Cross Network

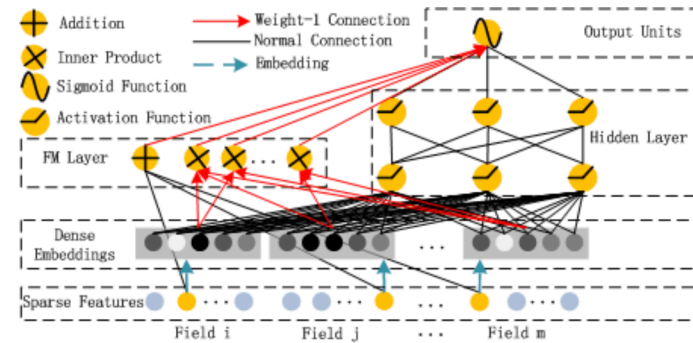
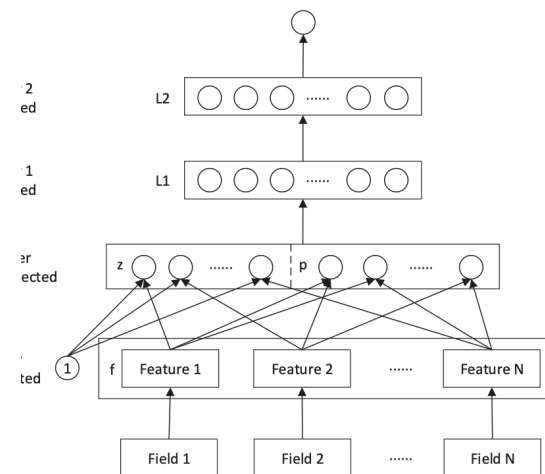
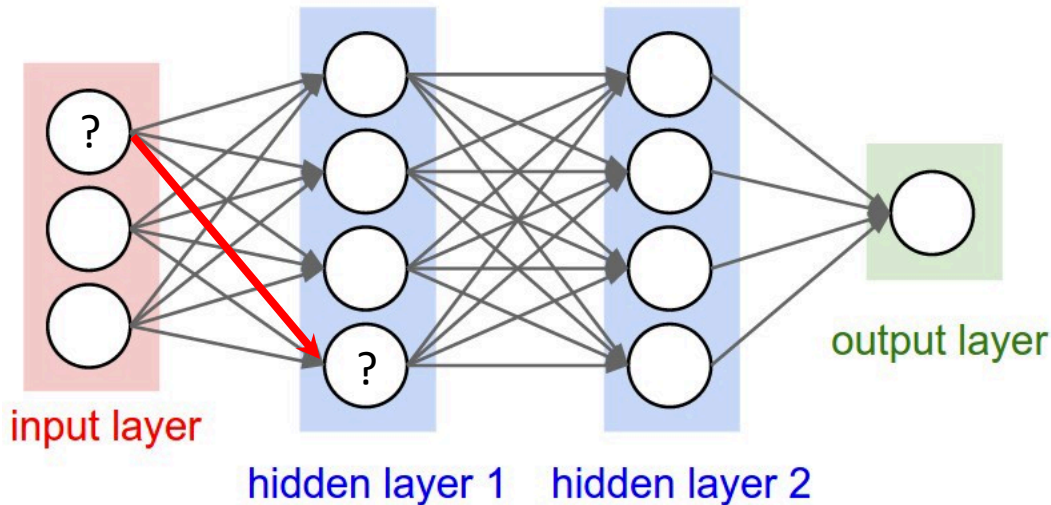


Figure 1: Wide & deep architecture of DeepFM. The wide and deep component share the same input raw feature vector, which enables DeepFM to learn low- and high-order feature interactions simultaneously from the input raw features.



Product-based Neural Network Architecture.

Deep neural network (DNN) module



- DNN
 - Widely used in CTR models
 - Unjustifiable element-wise computation within representations of input or hidden features
 - Unaffordable complexity for big feature dim or size

Image: <https://hackernoon.com/challenges-in-deep-learning-57bbf6e73bb>

Concerns of DNN

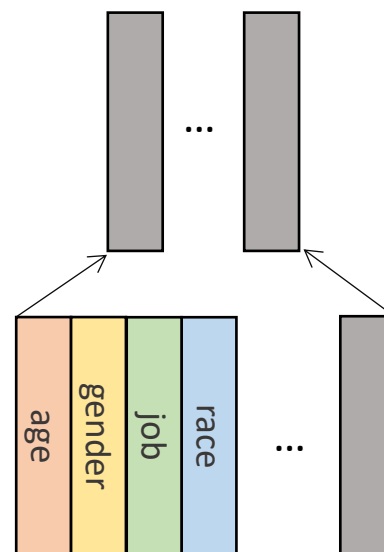
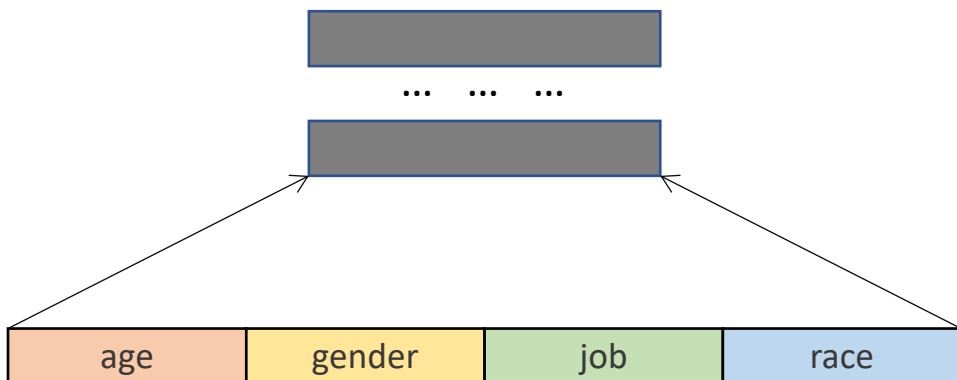
- Okay for online shopping with general purposes
 - Shopping on Amazon ...
- *NOT* okay for:
 - Medicine recommendation
 - Financial service recommendation
- Criteo:
 - 4 billions clicks in 24 hrs



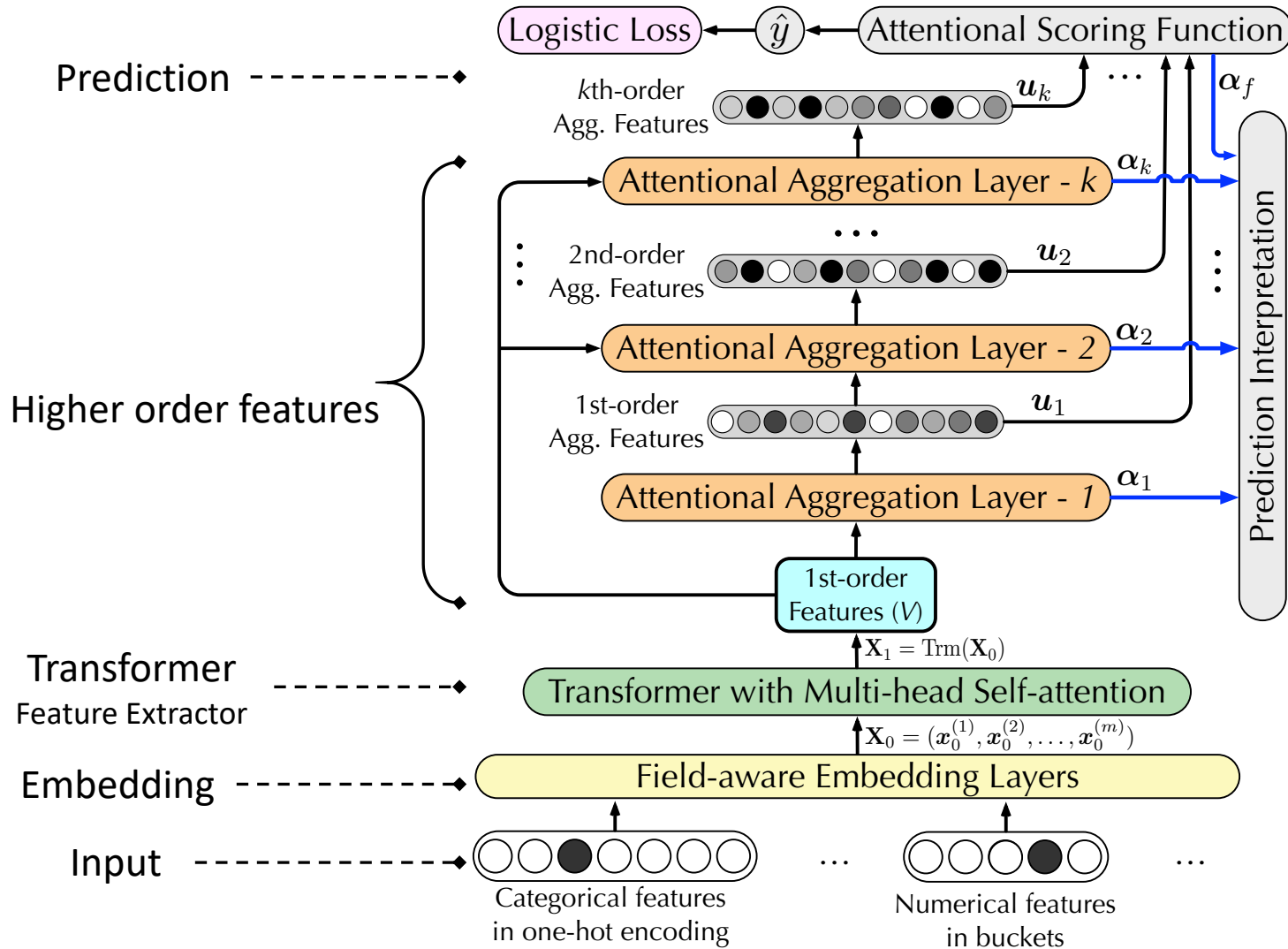
Images: <https://www.fajarmag.com/an-increasing-trend-of-online-shopping/>;
[https://www.tes.com/lessons/arP_sMT1GxDxHQ/medicine-by-the-minute](https://www.tes.com/lessons/arP_sMT1GxDxHQ/medicine-by-the-minute;);
<https://www.wealthandfinance-news.com/awards/>

Our idea -- InterHAt

- Interpretability
 - Attention mechanism
 - Avoid flat concatenation of features
 - Avoid DNN and dim-wise computation
- Efficiency
 - Shrunk problem size

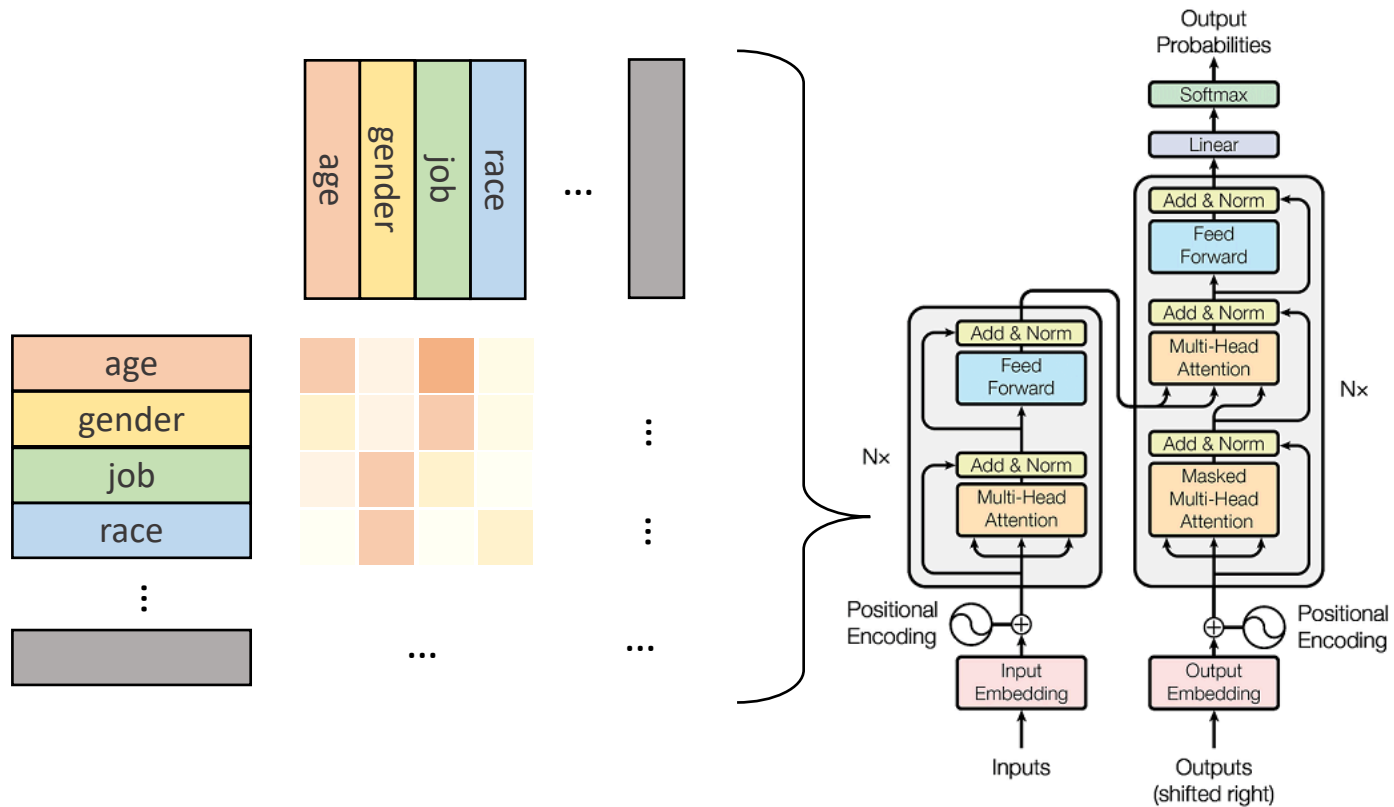


Interpretable CTR via Hierarchical Attention



Polysemy

○ Self attention in Transformer



Right Figure: Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems. 2017.

Hierarchical Attention

- Input: i-th order features:
- Generate aggregated feature:

$$\alpha_i^{(j)} = \frac{\exp(\mathbf{c}_i^T \text{ReLU}(\mathbf{W}_i \mathbf{x}_i^{(j)}))}{\sum_{j' \in F} \exp(\mathbf{c}_i^T \text{ReLU}(\mathbf{W}_i \mathbf{x}_i^{(j')}))},$$

$$\mathbf{u}_i = \text{AttentionalAgg}(\mathbf{X}_i) = \sum_{j=1}^m \alpha_i^{(j)} \mathbf{x}_i^{(j)},$$

- Output (i+1)-t order features:

$$\mathbf{x}_{i+1}^{(j)} = \mathbf{u}_i \circ \mathbf{x}_1^{(j)} + \mathbf{x}_i^{(j)}, \quad j \in \{1, \dots, m\},$$

Evaluation

- Datasets

- Performance evaluation

- Critio, Avazu, Frappe

- Interpretability study

- Movielens-1m dataset (reviews as clicks)

Dataset	Criteo	Avazu	Frappe
#. of features (C + N)	22 + 14	21 + 0	7 + 0
#. of total records	13.8M	12.1M	288K
#. of distinct features	605.7K	23.8K	5,382

- Baselines

- FM, Wide&Deep, DCN, PNN, DeepFM, xDeepFM

- Metrics

- Area Under ROC Curve (AUC)
 - Cross Entropy (LogLoss)

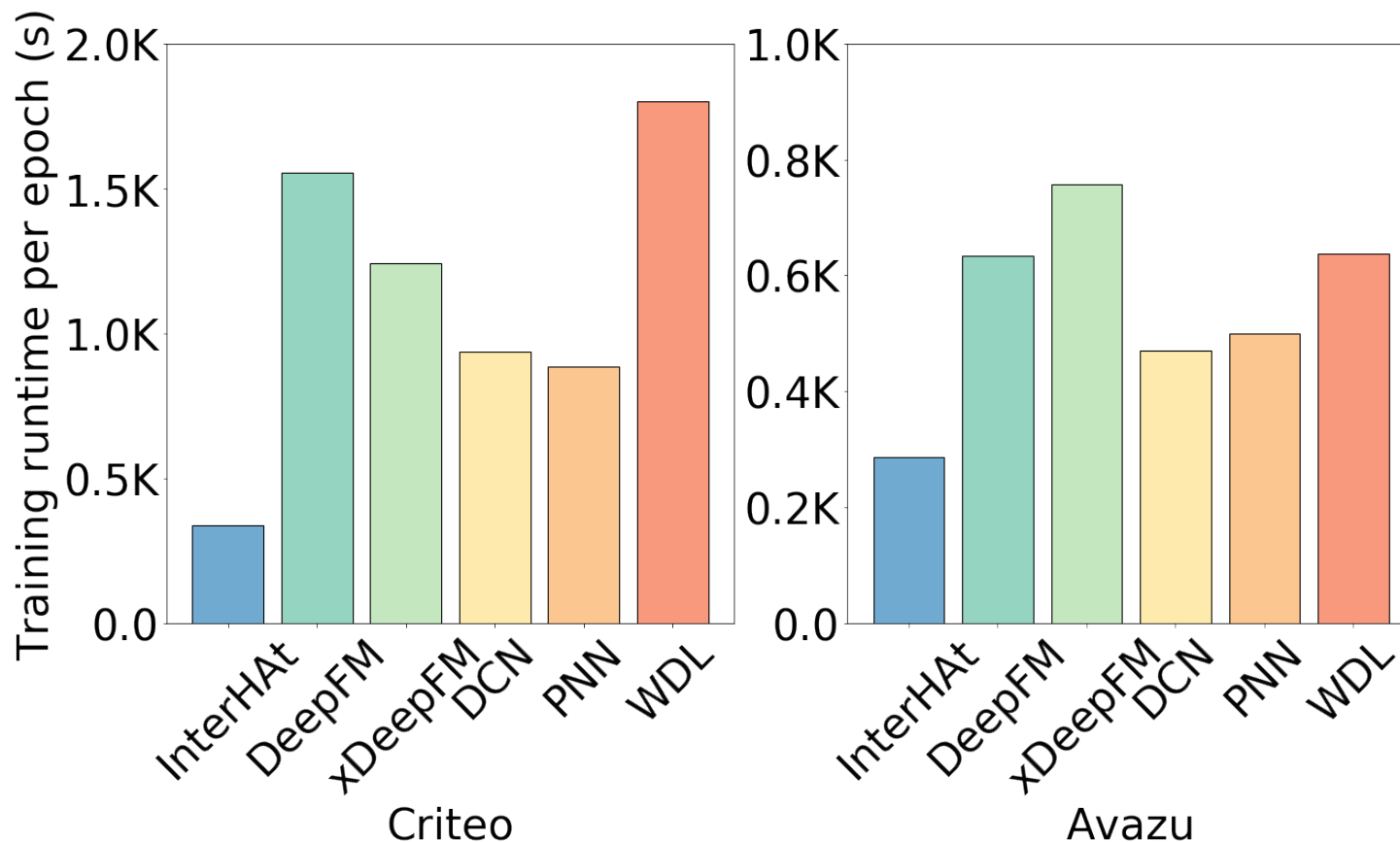
Performance

- Comparable with SOTA models
- Perform better on categorical features
 - SOTA models have close performance
 - Need better ways for encoding numeric features

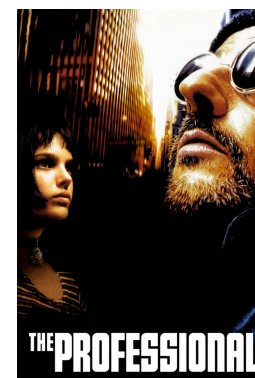
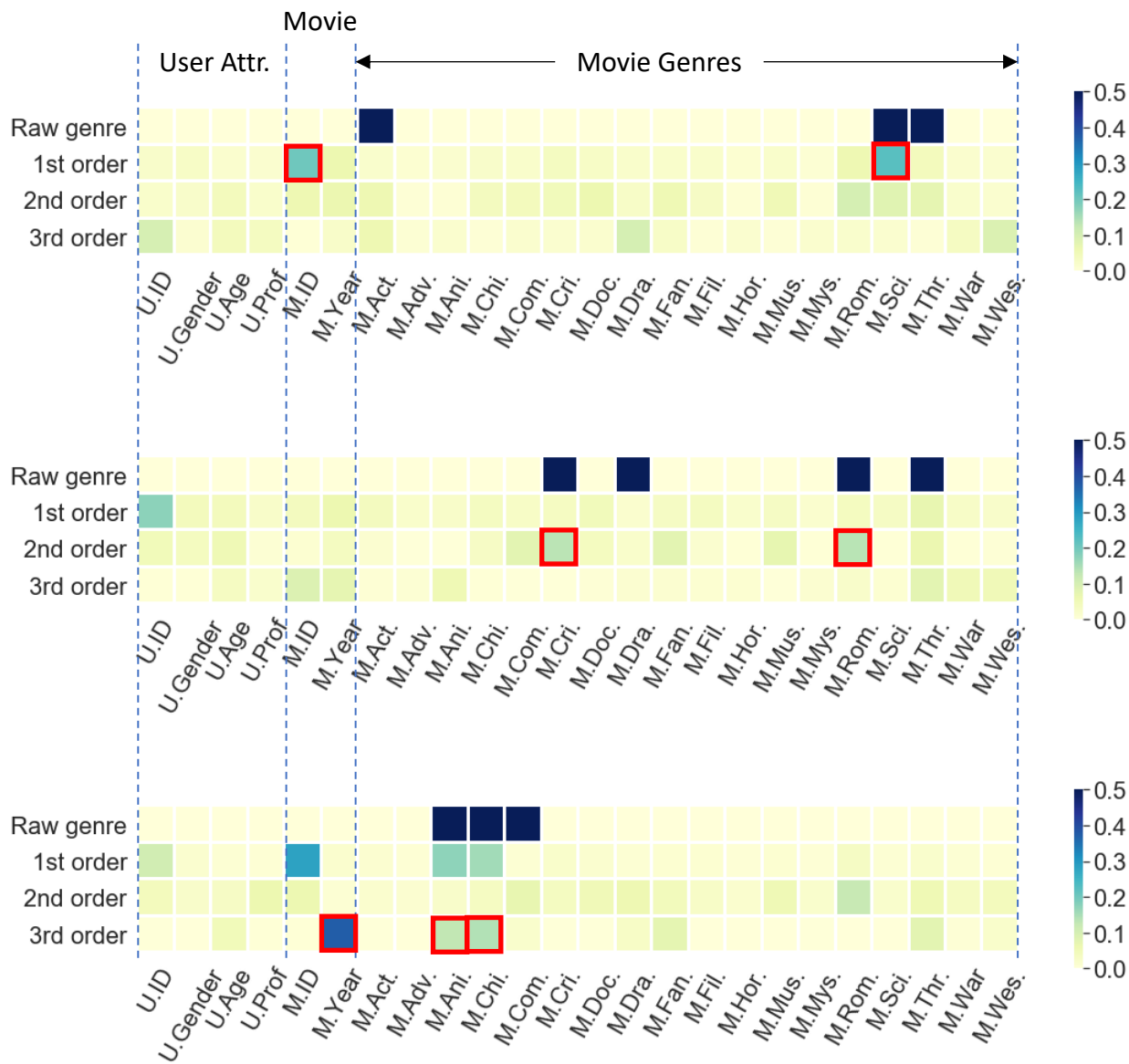
Dataset	Criteo		Avazu		Frappe	
Metrics	Logloss	AUC	Logloss	AUC	Logloss	AUC
FM	0.4814	0.7525	0.3951	0.7508	0.4480	0.8625
Wide&Deep	0.4577	0.7845	0.3920	0.7564	0.2571	0.9500
DCN	0.4590	0.7826	0.3921	0.7564	0.2335	0.9616
PNN	0.4547	0.7887	0.3916	0.7569	0.2177	0.9642
DeepFM	0.4560	0.7866	0.3920	0.7561	0.2410	0.9520
xDeepFM	0.4563	0.7874	0.3917	0.7569	0.2043	0.9694
InterHAt-S	0.4608	0.7820	0.3919	0.7577	0.2151	0.9616
InterHAt	0.4577	0.7845	0.3910	0.7582	0.2026	0.9696

Efficiency

- InterHAT trains faster than other baselines



Interpretability



Conclusion

- InterHAt:
 - Efficiency and interpretability issues of CTR task
 - Efficiency:
 - Avoiding **deep** fully connect neural networks
 - Interpretability:
 - Attention mechanism
 - Interpretability v.s. Explanability
 - Nice performances on both aspects!
 - Try it out:
 - <https://github.com/zyli93/InterHAt>

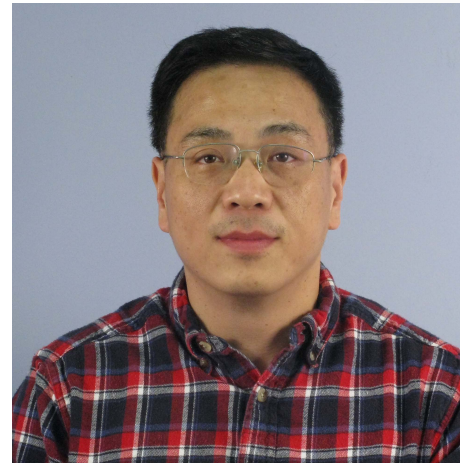
Questions?



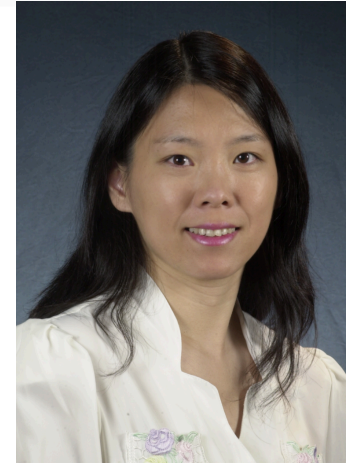
Wei Cheng



Yang Chen



Haifeng Chen



Wei Wang