DataFunSummit

From Feature Interaction to Sample Interaction: the new Trend for Deep CTR Models

从特征交互到数据交互：深度点击率模型的新趋势



张伟楠 上海交通大学
http://wnzhang.net

An Display Ad Example

Click or not?

大陆



河南省公安厅彻查“封丘36人入警35人身份不合规”

中封丘县公安局的36名受训人员，35人是公安局内部的文职或临时人员，与“民警必须具备公务员身份”的国家规定不符，引发该局内部

- 上海至成都沿江高铁提上日程 串联长江沿线22城市
- 2016号歼-20原型机曝光 已滑行测试(图)
- 日媒：中国或派万吨海警船巡钓鱼岛 打消耗战
- 外媒：中国开始研制隐身武装直升机 预计2020年交付
- 习近平关于中美关系的十个判断
- 住建部黑臭水沟整治工作指南：9成百姓满意才能达标
- 陕西：职校“校长”让女学生陪酒 学校被撤除
- 揭秘“团团伙伙”的武钢漩涡和落马高管

国际



巴塞罗那200万人游行 呼吁加泰罗尼亚独立(图)

- 李炜光：收税是不公平的恶？
- 许章润：超级大国没有纯粹内政
- 刘昀献:国外政党联系群众的路径研究

时局观



民革中央副主席：中共从未否定国民党抗战作用

- 施芝鸿：文革基础上搞改革致一个时期市场官场乱象
- 朱维群回应争议：尊重民族差异而不强化
- 伊协副会长：穆斯林不应因宗教功修忽视社会责任

领袖圈



奥巴马54岁啦，当7年总统人苍老了头发也



海绵城市 未来之城
水危机：青岛告急
探访中国绿化博览会
帝都吸引华人首富
凤凰房产 诚邀加盟

谈华山论剑与中国精神
黑龙江创新驱动三步棋
《印记》之江城夜未眠
办公环境搜查令
圈层生活尽在凤凰会

精彩视频

凤凰联播台



菲媒曝菲律宾军演针对中国 直指南海生命线

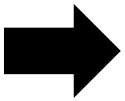
播放数：2602282

[http://news.ifeng.com]

User Response Estimation Problem

- Click-through rate estimation as an example

- Date: 20160320
- Hour: 14
- Weekday: 7
- IP: 119.163.222.*
- Region: England
- City: London
- Country: UK
- Ad Exchange: Google
- Domain: yahoo.co.uk
- URL: http://www.yahoo.co.uk/abc/xyz.html
- OS: Windows
- Browser: Chrome
- Ad size: 300*250
- Ad ID: a1890
- User tags: Sports, Electronics



Click (1) or not (0)?

Predicted CTR (0.15)

CTR and Data

- Click-Through Rate (CTR) Estimation is an essential task
  - It predicts the probability of a user clicking an item based on the logged behavior data
  - Applications: e-commerce, online advertising, recommender systems
- Data format: multi-categorical data with binary labels

USER ID	CITY	OCCUPATION	ITEM ID	CATEGORY	BRAND	CLICK
U1	LA	student	I1	T-shirt	B1	1
U2	NYC	student	I1	T-shirt	B1	1
U1	LA	student	I2	cell phone	B2	0
U3	London	manager	I3	cell phone	B3	0

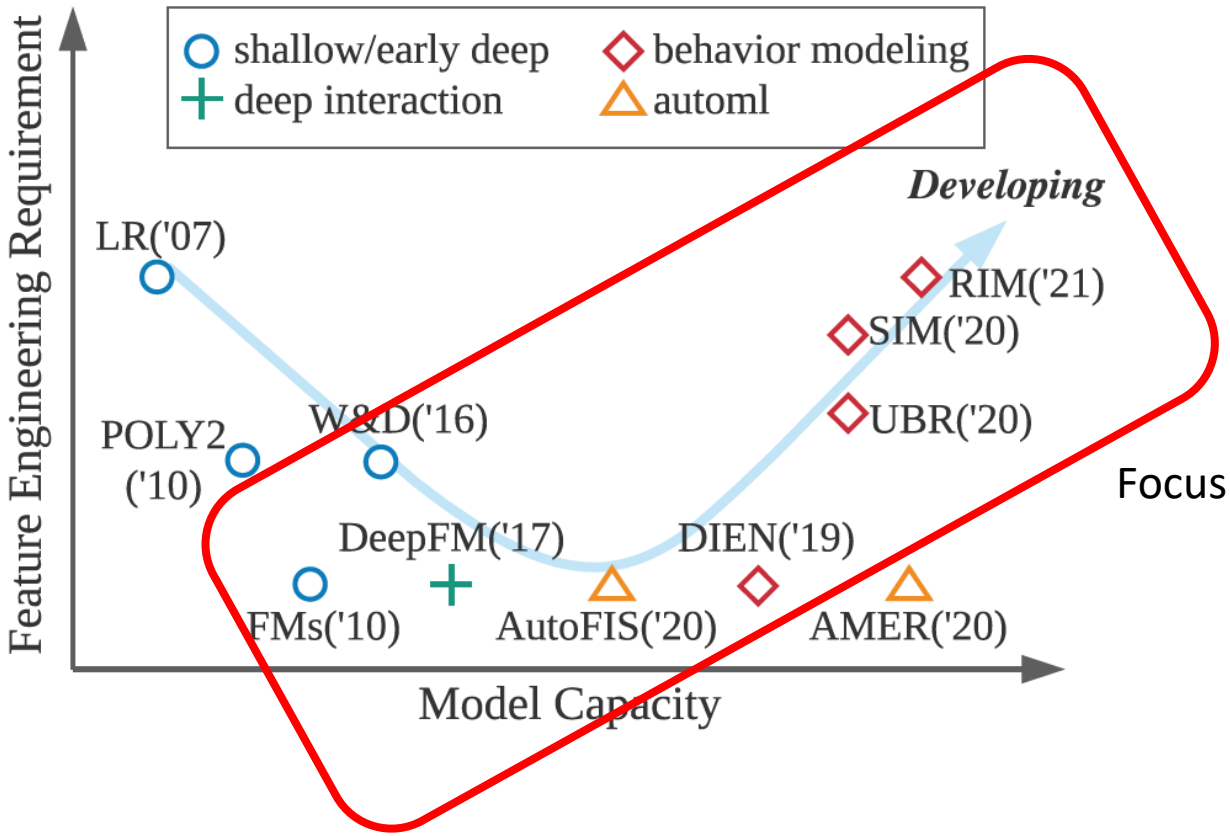
- CTR estimation as a binary classification formulation

$$l(x, y, \Theta) = -y \log \sigma(f_\Theta(x)) - (1 - y) \log(1 - \sigma(f_\Theta(x)))$$

Weinan Zhang, Jiarui Qin, Wei Guo, Ruiming Tang, Xiuqiang He. Deep Learning for Click-Through Rate Estimation. IJCAI 2021 (Survey Track)

CTR Models (2007 to 2021)

- Developing trend of CTR estimation models
  - Feature engineering requirement
  - Model capacity

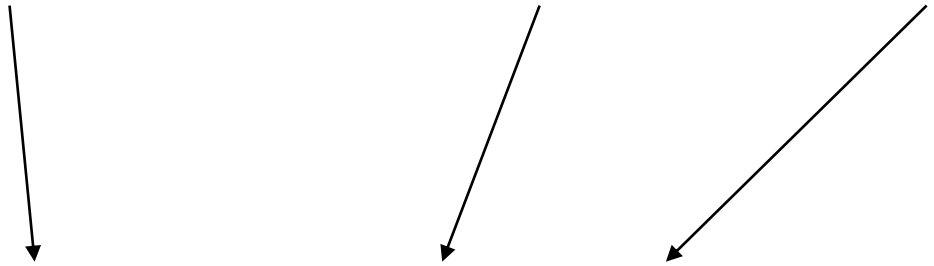


Weinan Zhang, Jiarui Qin, Wei Guo, Ruiming Tang, Xiuqiang He. Deep Learning for Click-Through Rate Estimation. IJCAI 2021 (Survey Track)

Feature Representation

- Binary one-hot encoding of categorical data

x=[Weekday=Wednesday, Gender=Male, City=London]



x=[0,0,1,0,0,0,0 0,1 0,0,1,0…0]

High dimensional sparse binary feature vector

- Label representation

y = 1 (click) or 0 (non-click)

Logistic Regression

- Prediction

$$\hat{y} = \frac{1}{1 + e^{-\boldsymbol{w}^T \boldsymbol{x}}}$$

- Cross Entropy Loss

$$\mathcal{L}(y, \hat{y}) = -y \log \hat{y} - (1 - y) \log(1 - \hat{y})$$

- Stochastic Gradient Descent Learning

$$\boldsymbol{w} \leftarrow (1 - \lambda)\boldsymbol{w} + \eta(y - \hat{y})\boldsymbol{x}$$

Lee et al. Estimating Conversion Rate in Display Advertising from Past Performance Data. KDD 12

Factorization Machines

- Prediction based on feature embedding

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

Logistic Regression — Feature Interactions

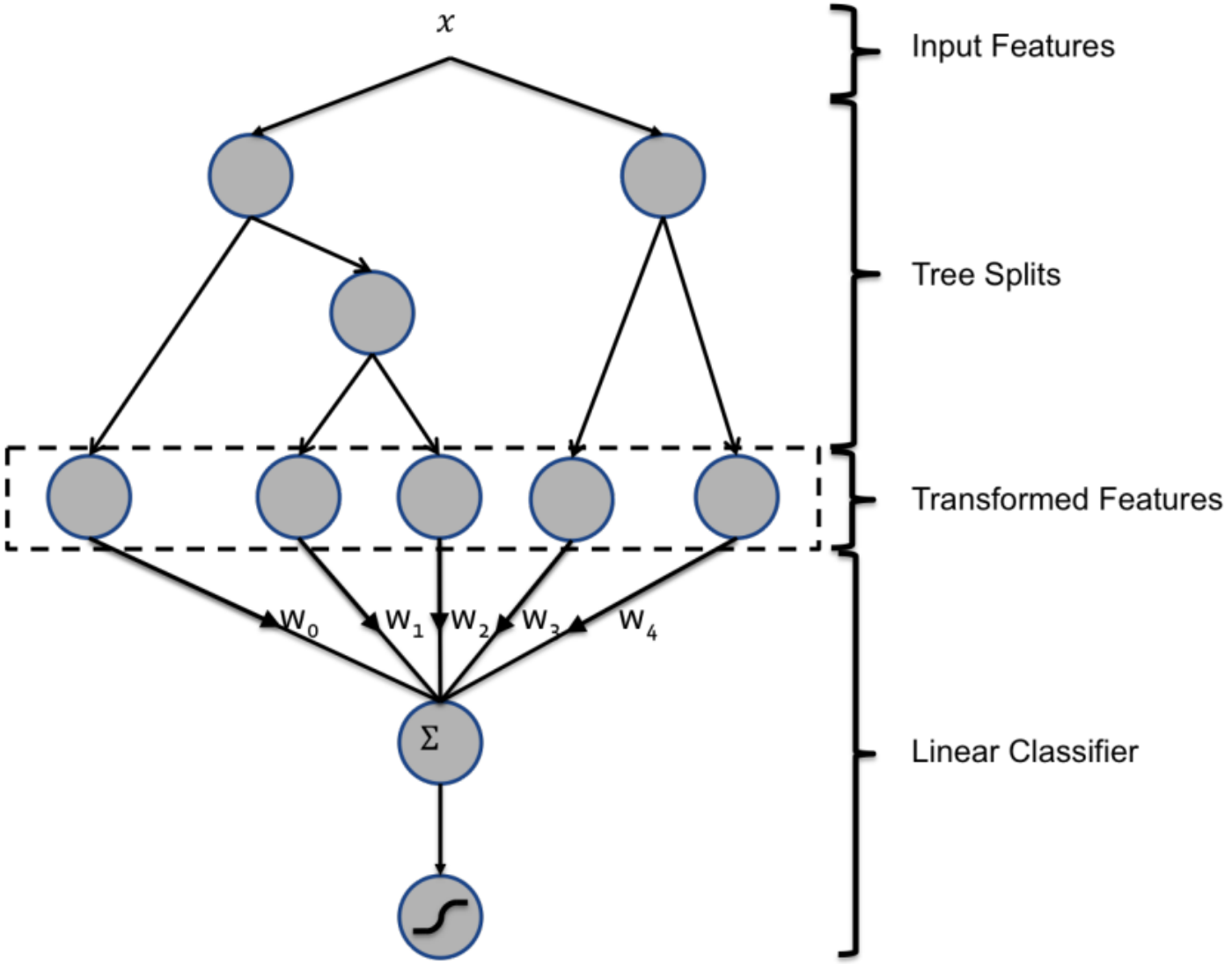
For x=[Weekday=Friday, Gender=Male, City=Shanghai]

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

Rendle. Factorization machines. ICDM 2010.

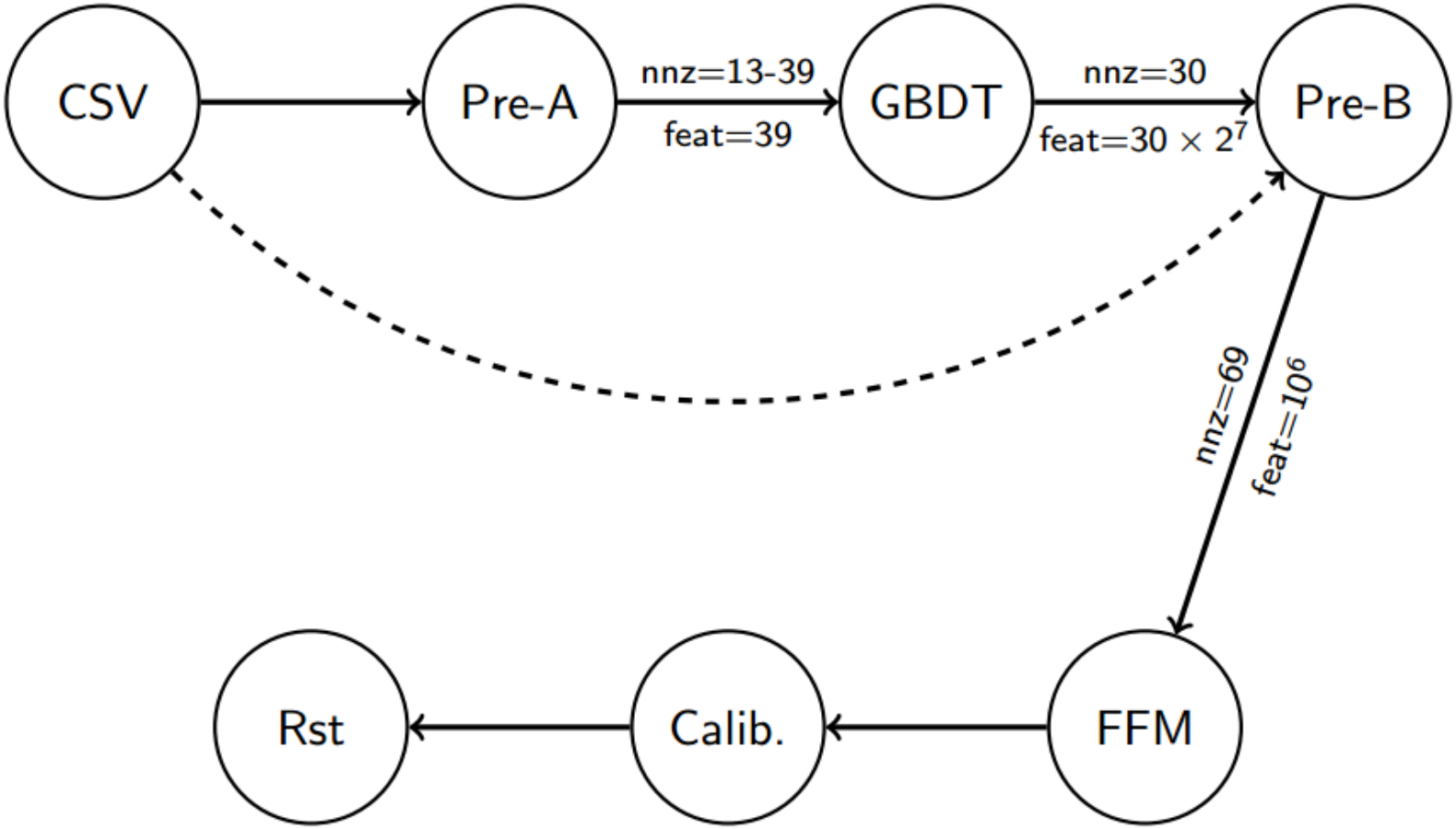
Oentaryo et al. Predicting response in mobile advertising with hierarchical importance-aware factorization machine. WSDM 14.

Combined Models: GBDT + LR



He et al. Practical Lessons from Predicting Clicks on Ads at Facebook . ADKDD 2014.

Combined Models: GBDT + FM

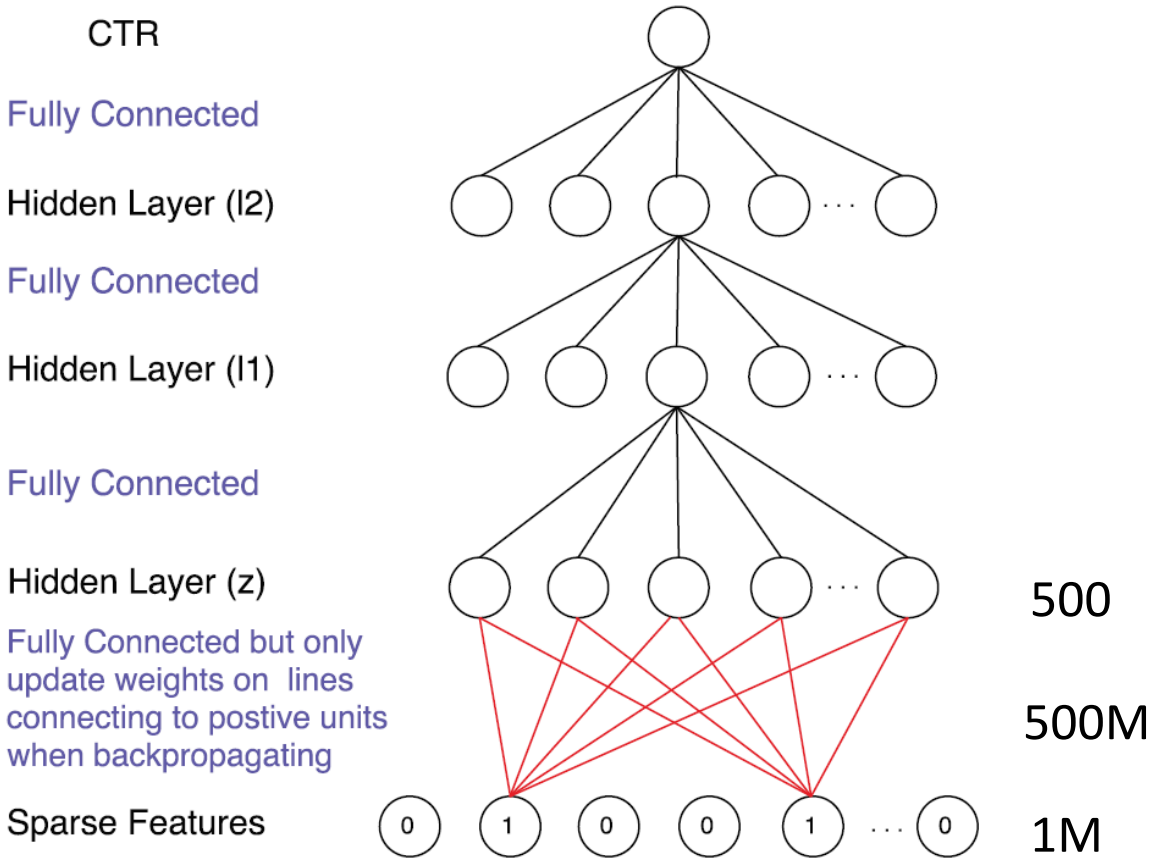


"nnz" means the number of non-zero elements of each impression; "feat" represents the size of feature space.

http://www.csie.ntu.edu.tw/~r01922136/kaggle-2014-criteo.pdf

Neural Network Models

- Difficulty: Impossible to directly deploy neural network models on such data



E.g., input features 1M, first layer 500, then 500M parameters for first layer

Review Factorization Machines

- Prediction based on feature embedding

$$y_{\mathrm{FM}}(\boldsymbol{x}) := \mathrm{sigmoid}\left(w_0 + \sum_{i=1}^{N} w_i x_i + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle x_i x_j\right)$$

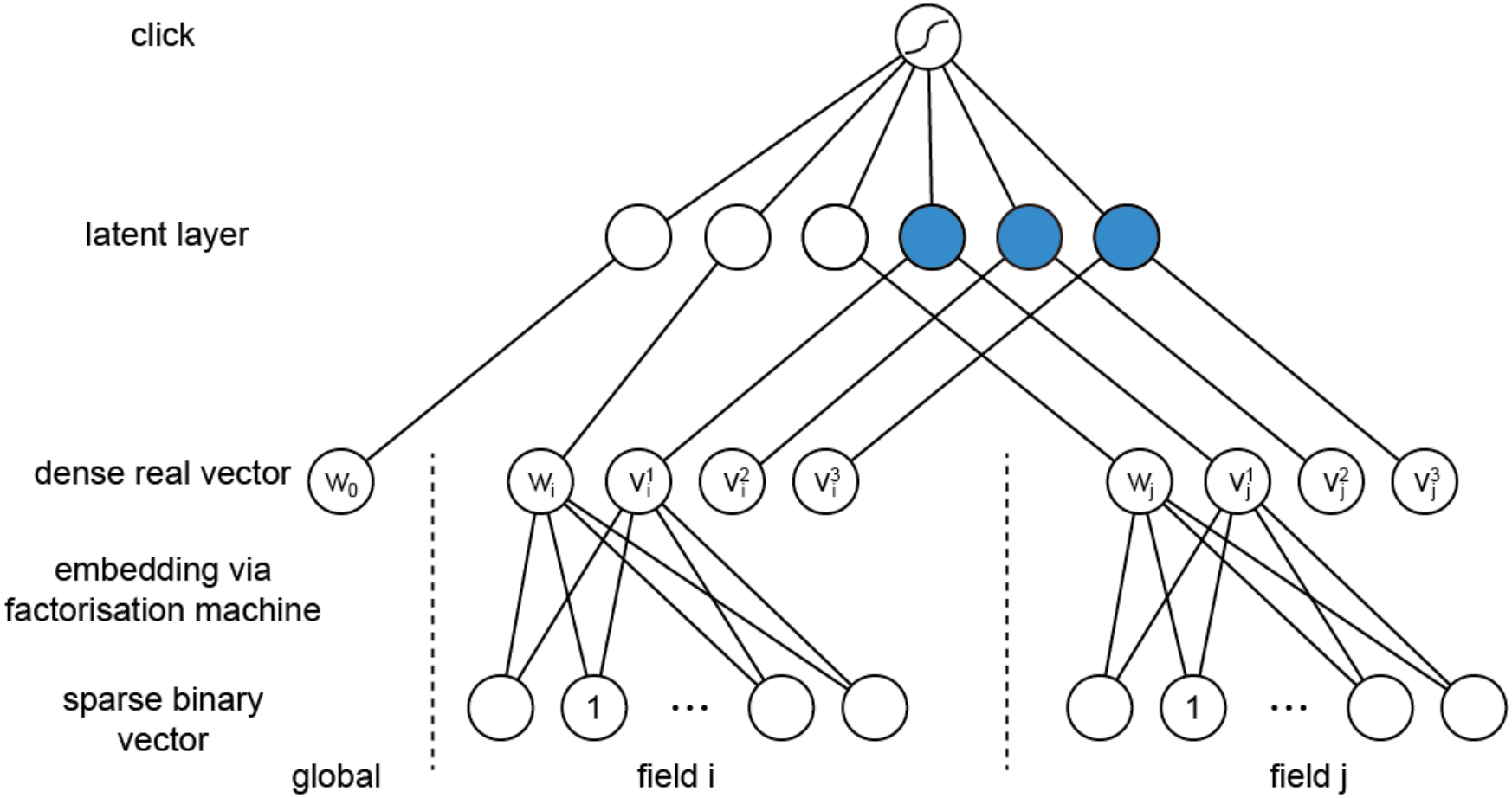
Logistic Regression Feature Interactions

  - Embed features into a k-dimensional latent space
  - Explore the feature interaction patterns using vector inner-product

Rendle. Factorization machines. ICDM 2010.

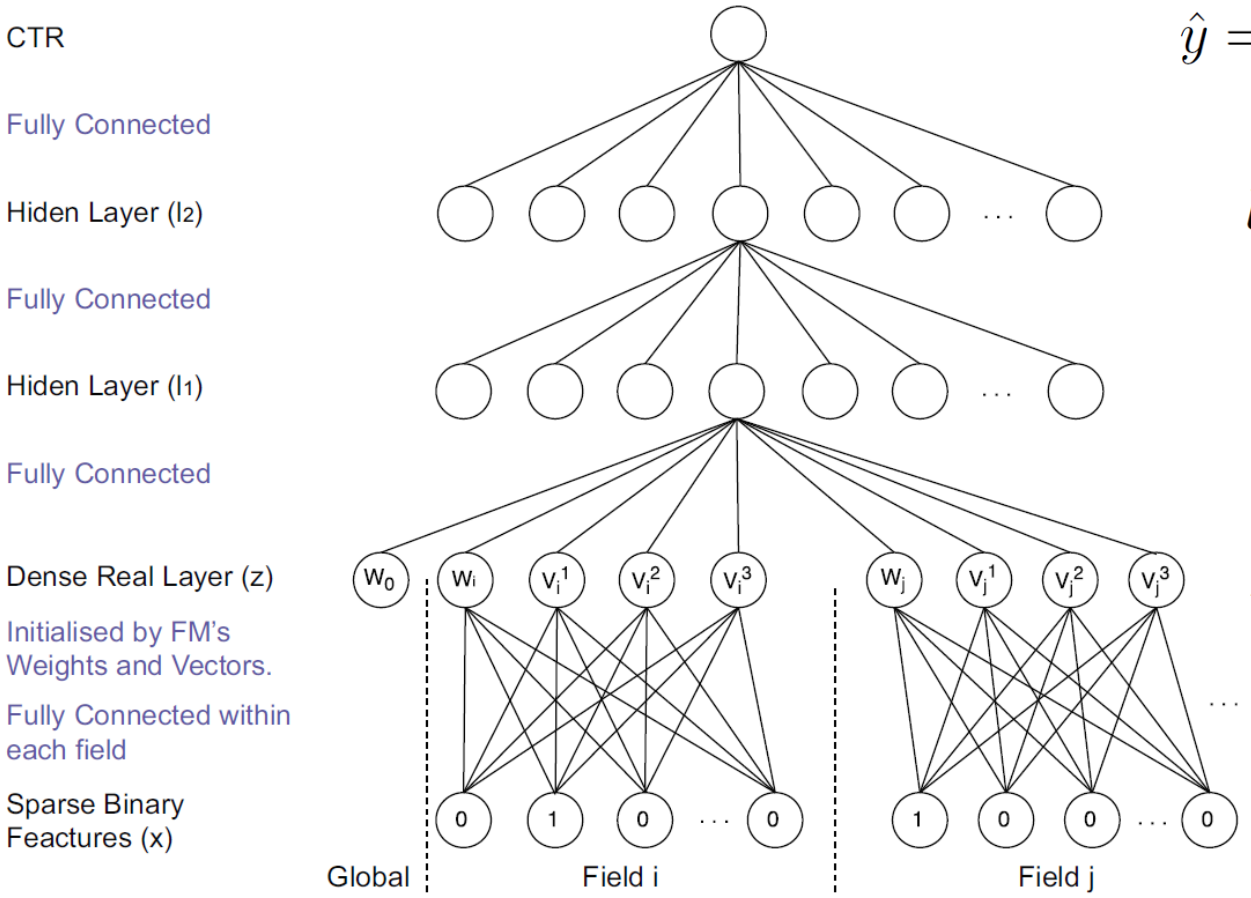
Oentaryo et al. Predicting response in mobile advertising with hierarchical importance-aware factorization machine. WSDM 14.

Factorization Machine is a Neural Network



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

Factorization-machine supported Neural Networks (FNN)



$$\hat{y} = \text{sigmoid}(\boldsymbol{W}_3 \boldsymbol{l}_2 + b_3)$$

$$\boldsymbol{l}_2 = \tanh(\boldsymbol{W}_2 \boldsymbol{l}_1 + \boldsymbol{b}_2)$$

$$\boldsymbol{l}_1 = \tanh(\boldsymbol{W}_1 \boldsymbol{z} + \boldsymbol{b}_1)$$

$$\boldsymbol{z}_i = (w_i, v_i^1, v_i^2, \ldots, v_i^K)$$
$$= \boldsymbol{W}_0^i \cdot \boldsymbol{x}[\text{start}_i : \text{end}_i]$$

[Factorisation Machine Initialised]

Zhang et al. Deep Learning over Multi-field Categorical Data – A Case Study on User Response Prediction. ECIR 16.

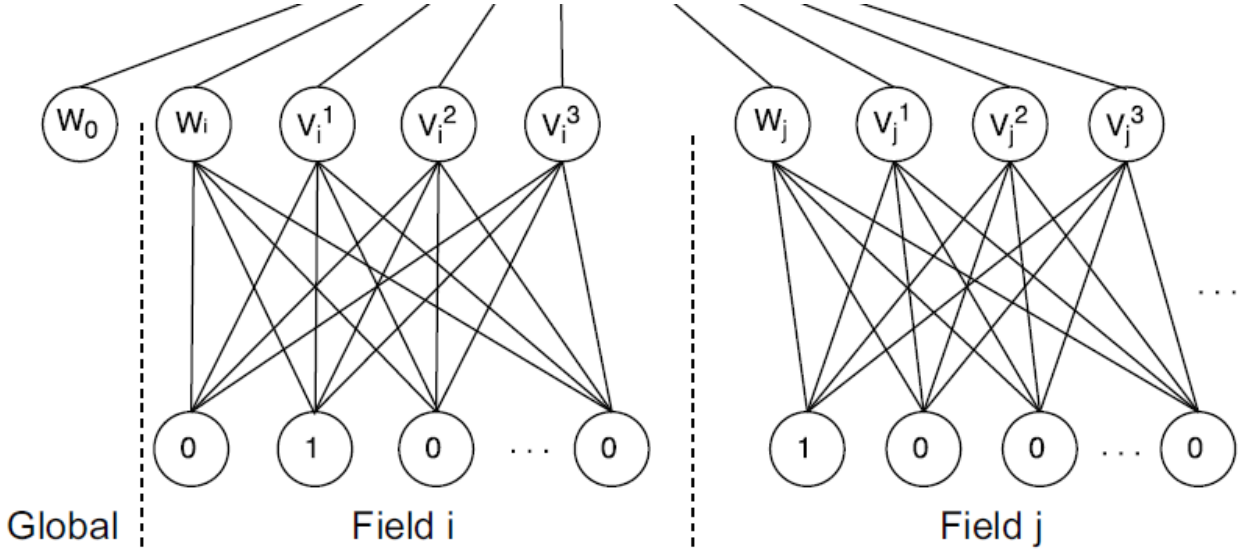
Factorization-machine supported Neural Networks (FNN)

Dense Real Layer (z)

Initialised by FM's Weights and Vectors.

Fully Connected within each field

Sparse Binary Feactures (x)



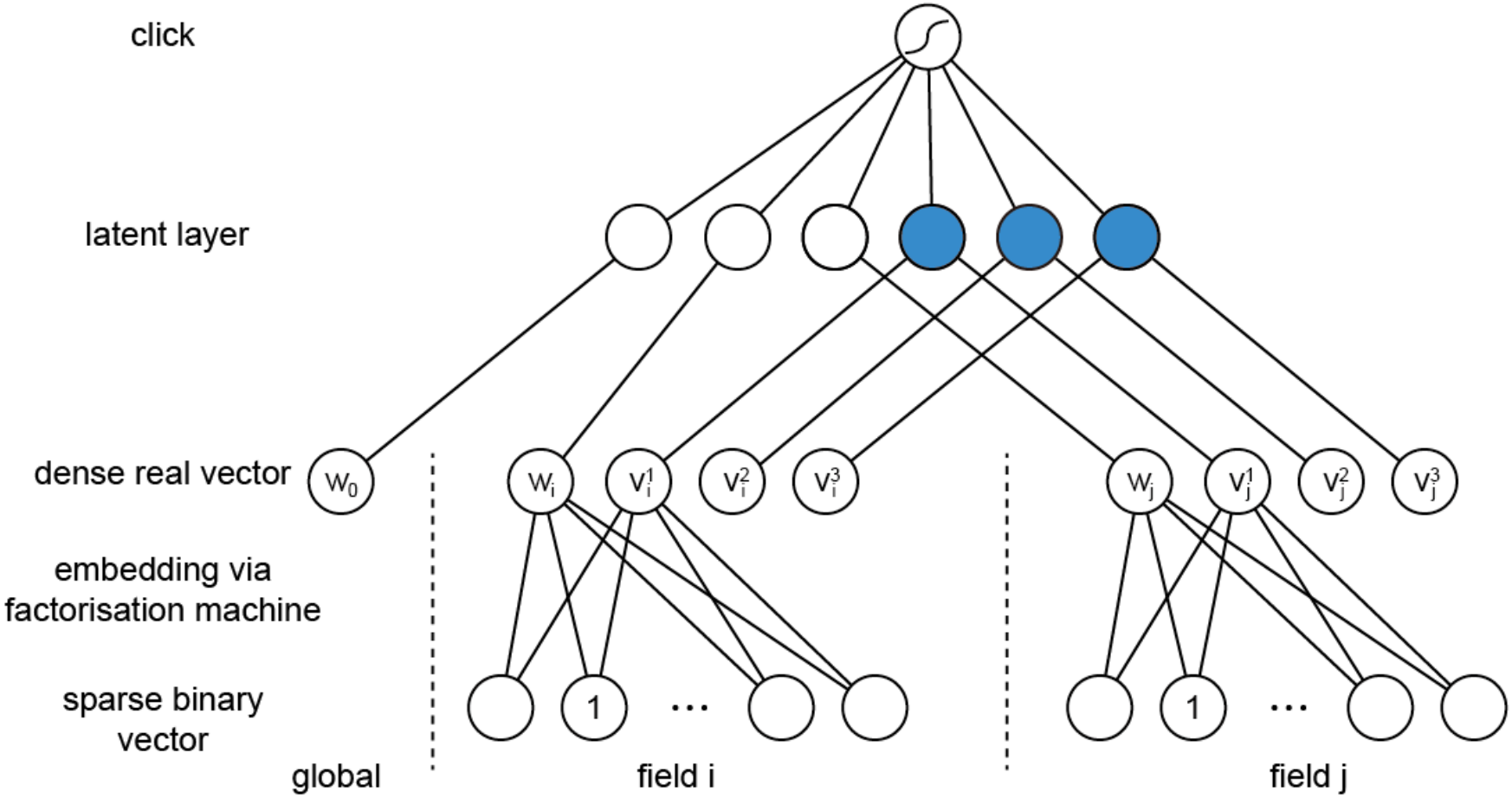
- Chain rule to update factorisation machine parameters

$$\frac{\partial L(y,\hat{y})}{\partial \boldsymbol{W}_0^i} = \frac{\partial L(y,\hat{y})}{\partial \boldsymbol{z}_i}\frac{\partial \boldsymbol{z}_i}{\partial \boldsymbol{W}_0^i} = \frac{\partial L(y,\hat{y})}{\partial \boldsymbol{z}_i}\boldsymbol{x}[\text{start}_i : \text{end}_i]$$

$$\boldsymbol{W}_0^i \leftarrow \boldsymbol{W}_0^i - \eta \cdot \frac{\partial L(y,\hat{y})}{\partial \boldsymbol{z}_i}\boldsymbol{x}[\text{start}_i : \text{end}_i].$$

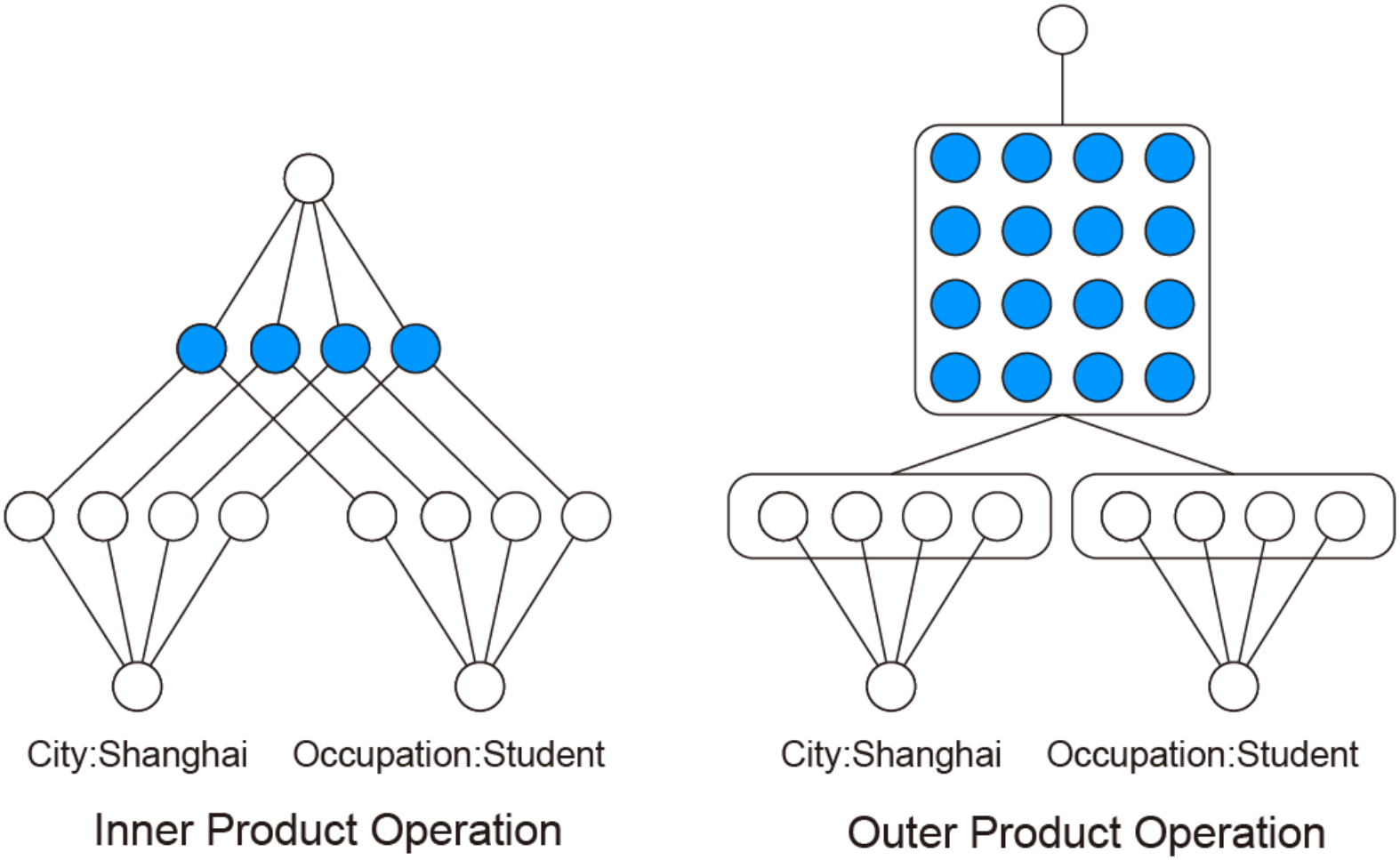
Zhang et al. Deep Learning over Multi-field Categorical Data – A Case Study on User Response Prediction. ECIR 16.

But factorization machine is still different from common additive neural networks



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

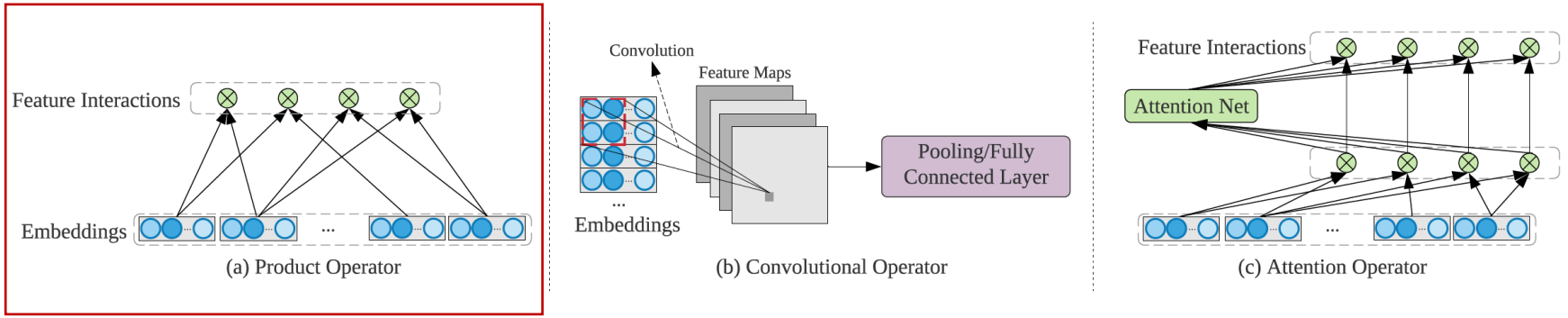
Product Operations as Feature Interactions



Yanru Qu et al. Product-based Neural Networks for User Response Prediction. ICDM 2016.

Feature Interaction

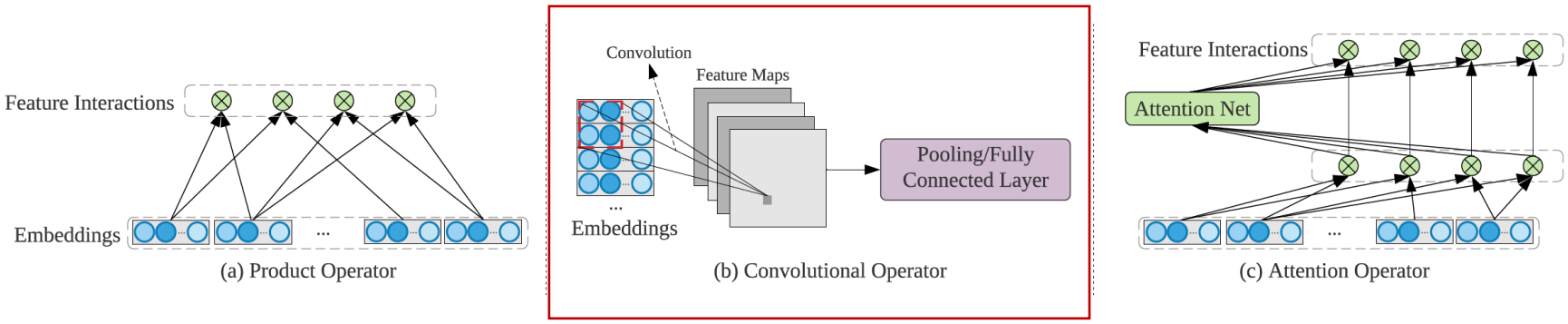
- Feature interaction operators
  - Product operator
    - PNN (IPNN, OPNN) [1], NFM [2], DeepFM[: inner and outer product
    - PIN [3]: micro neural network as feature interaction operator
    - DCN [4] (DCN V2 [5]), CIN [6]: cross network



[1] Qu et al. Product-based neural networks for user response prediction. ICDM, 2016.
[2] He et al. Neural factorization machines for sparse predictive analytics. SIGIR, 2017.
[3] Qu et al. Product-based neural networks for user response prediction over multi-field categorical data. TOIS, 2018.
[4] Wang et al. Deep & cross network for ad click predictions. ADKDD, 2017.
[5] Wang et al. Dcn-m: Improved deep & cross network for feature cross learning in web-scale learning to rank systems.
[6] Lian et al. xdeepfm: Combining explicit and implicit feature interactions for recommender systems. KDD 2018.

Feature Interaction

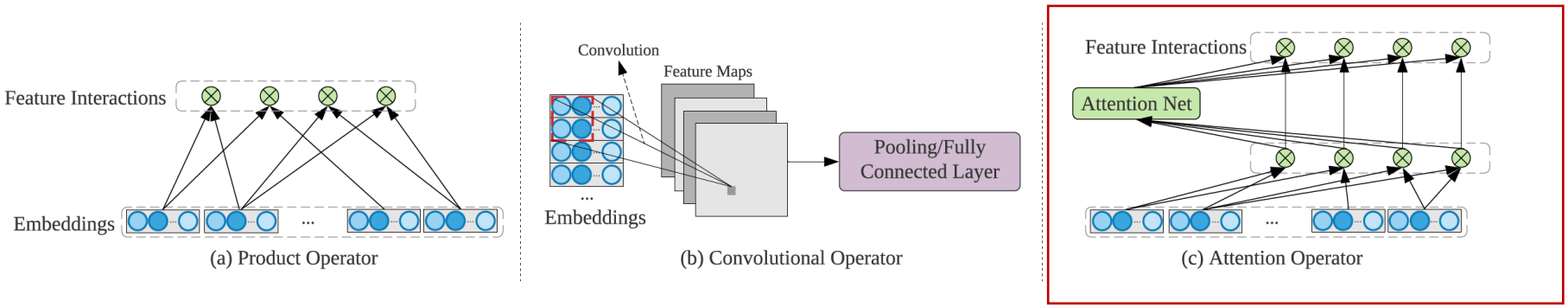
- Feature interaction operators
  - Convolutional operators
    - Use convolution networks to model the feature interaction on the feature map
    - CCPM [1], FGCNN [2]
    - GCN: FiGNN [3]



[1] Liu et al. A convolutional click prediction model. CIKM, 2015.
[2] Liu et al. Feature generation by convolutional neural network for click-through rate prediction. WWW 2019.
[3] Li et al. Fi-gnn: Modeling feature interactions via graph neural networks for ctr prediction. CIKM, 2019.

Feature Interaction

- Feature interaction operators
  - Attention operators
    - Use attention mechanism to weight different feature interactions: AFM [1], FiBiNet [2]
    - Self attention: AutoInt [3], InterHAt [4]



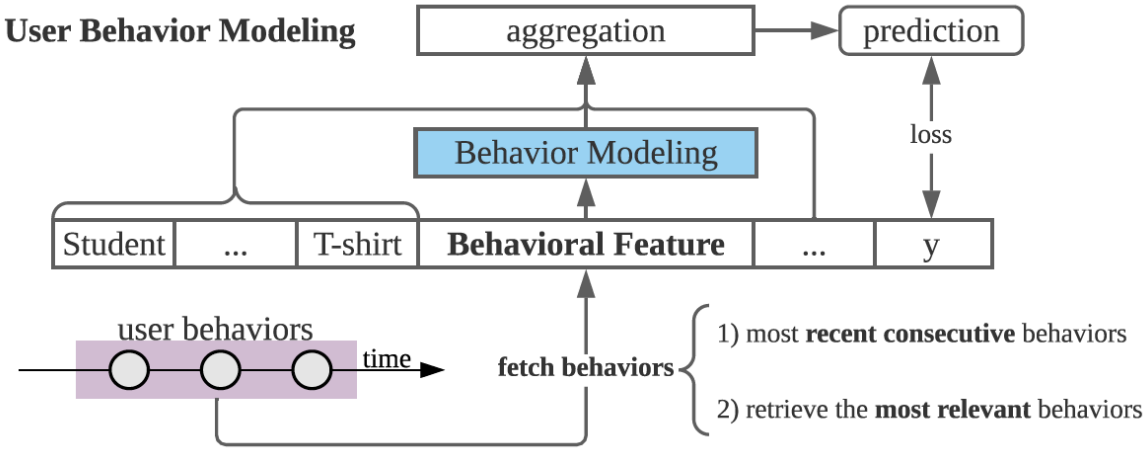
[1] Xiao et al. Learning the weight of feature interactions via attention networks. IJCAI, 2017.
[2] Huang et al. Fibinet: combining feature importance and bilinear feature interaction for click-through rate prediction. RecSys 2019.
[3] Song et al. Autoint: Automatic feature interaction learning via self attentive neural networks. CIKM, 2019.
[4] Li et al. Interpretable click-through rate prediction through hierarchical attention. WSDM 2020.

User Behavior Modeling

- Basic structure of user behavior modeling based CTR estimation
  - Behavior feature is a special multi-value feature consists of the item sequence one user has interacted with

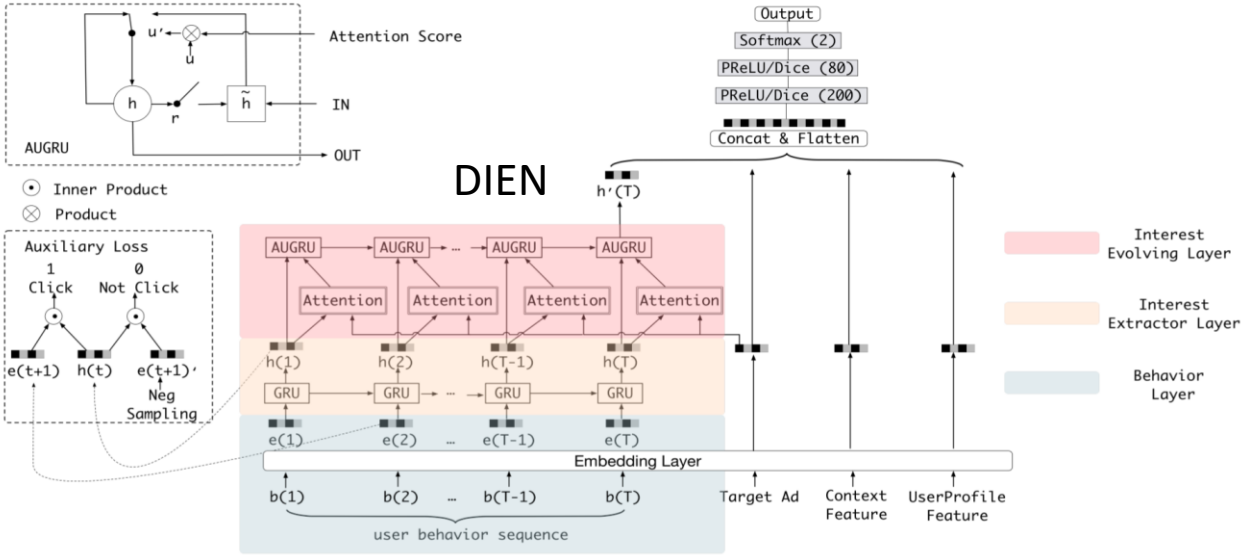
$$f_{\Theta}^{\text{UBM}}(x) = \text{MLP}_{\phi}([v_1, v_2, \ldots, v_{m-1}, \mathcal{U}_{\psi}(x_m)])$$

  - Model the temporal patterns inside the behavior sequences



User Behavior Modeling

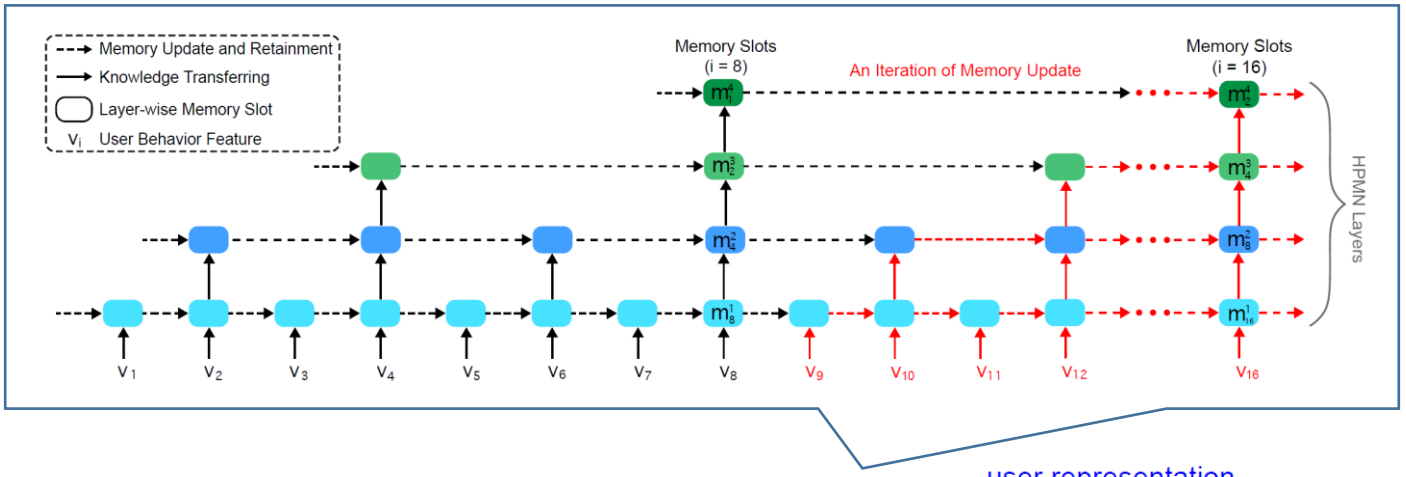
- Attention based models:
  - Attention is utilized to model the importance of different behaviors of user to the prediction target
  - DIN [1], DIEN [2] (GRU+Attention)
  - Self-attention: BST [3]; Session-split behavior model: DSIN [4]



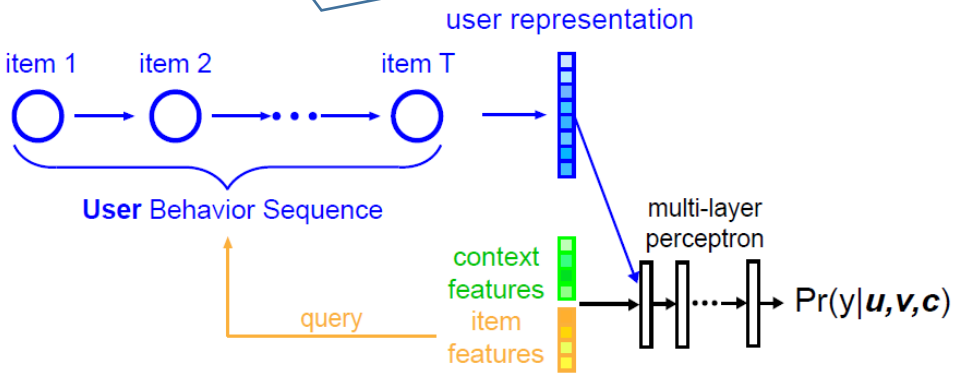
[1] Zhou et al. Deep interest network for clickthrough rate prediction. KDD 2018.
[2] Zhou et al. Deep interest evolution network for click-through rate prediction. AAAI 2019.
[3] Chen et al. Behavior sequence transformer for e-commerce recommendation in alibaba. DLP-KDD, 2019.
[4] Li et al. Deep session interest network for click-through rate prediction. IJCAI 2019.

User Behavior Modeling

- Memory network based models:



- Use memory network to store the user states and update the states when new behaviors happen
- Model very long user behavior sequence
- HPMN [1], MIMN [2]

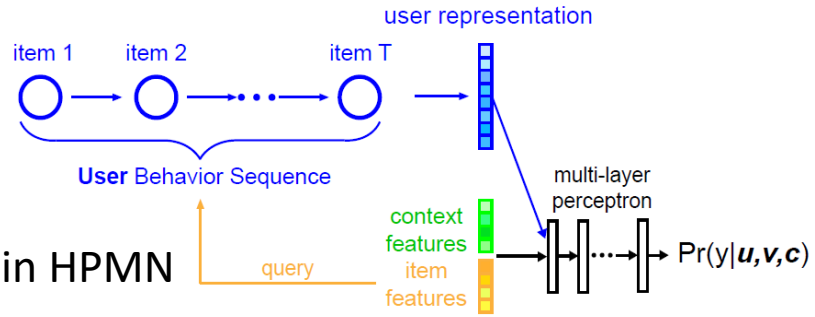


[1] Ren et al. Lifelong sequential modeling with personalized memorization for user response prediction. SIGIR 2019.
[2] Pi et al. Practice on long sequential user behavior modeling for click-through rate prediction. KDD 2019.

User Behavior Modeling

- Retrieval based models:
  - Use **retrieved relevant** behaviors instead of *recent consecutive* behaviors
  - Especially useful for utilizing very long user behavior sequences
  - Build index structure for retrieving behaviors
  - UBR [1], SIM [2], RIM [3]

As already hinted in HPMN



[1] Qin et al. User behavior retrieval for click-through rate prediction. SIGIR 2020.
[2] Pi et al. Search-based user interest modeling with lifelong sequential behavior data for click-through rate prediction. CIKM 2020.
[3] Qin et al. Retrieval & interaction machine for tabular data prediction. KDD 2021.

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：https://d.book118.com/446102031032010042