Deep Learning over Multi-field Categorical Data

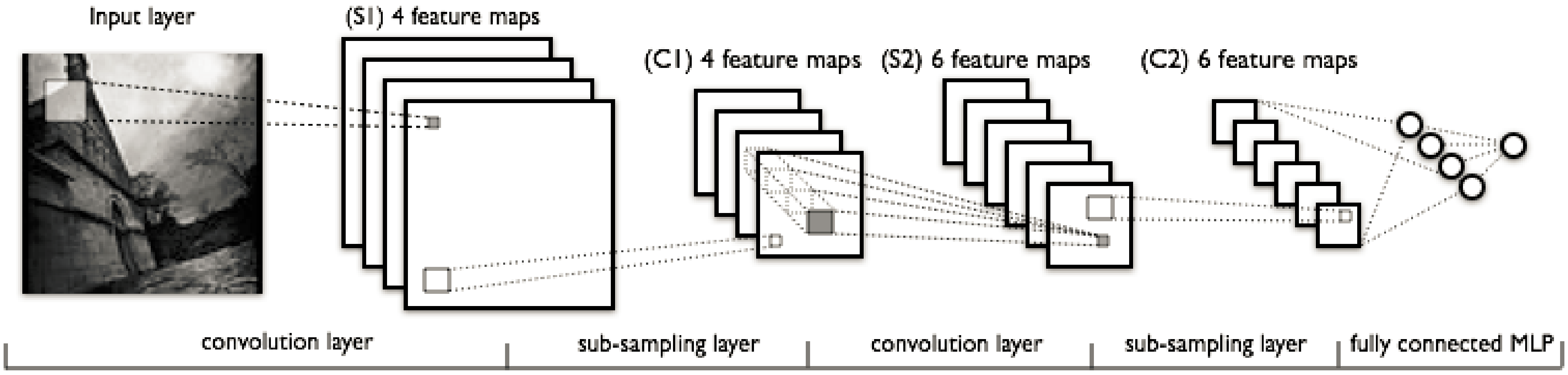
- A Case Study on User Response Prediction in Display Ads

Weinan Zhang
Shanghai Jiao Tong University & ulu.ai
w.zhang@cs.ucl.ac.uk

with colleagues Tianming Du, Jun Wang from University College London
and Yanru Qu, Han Cai and Kan Ren from Shanghai Jiao Tong University

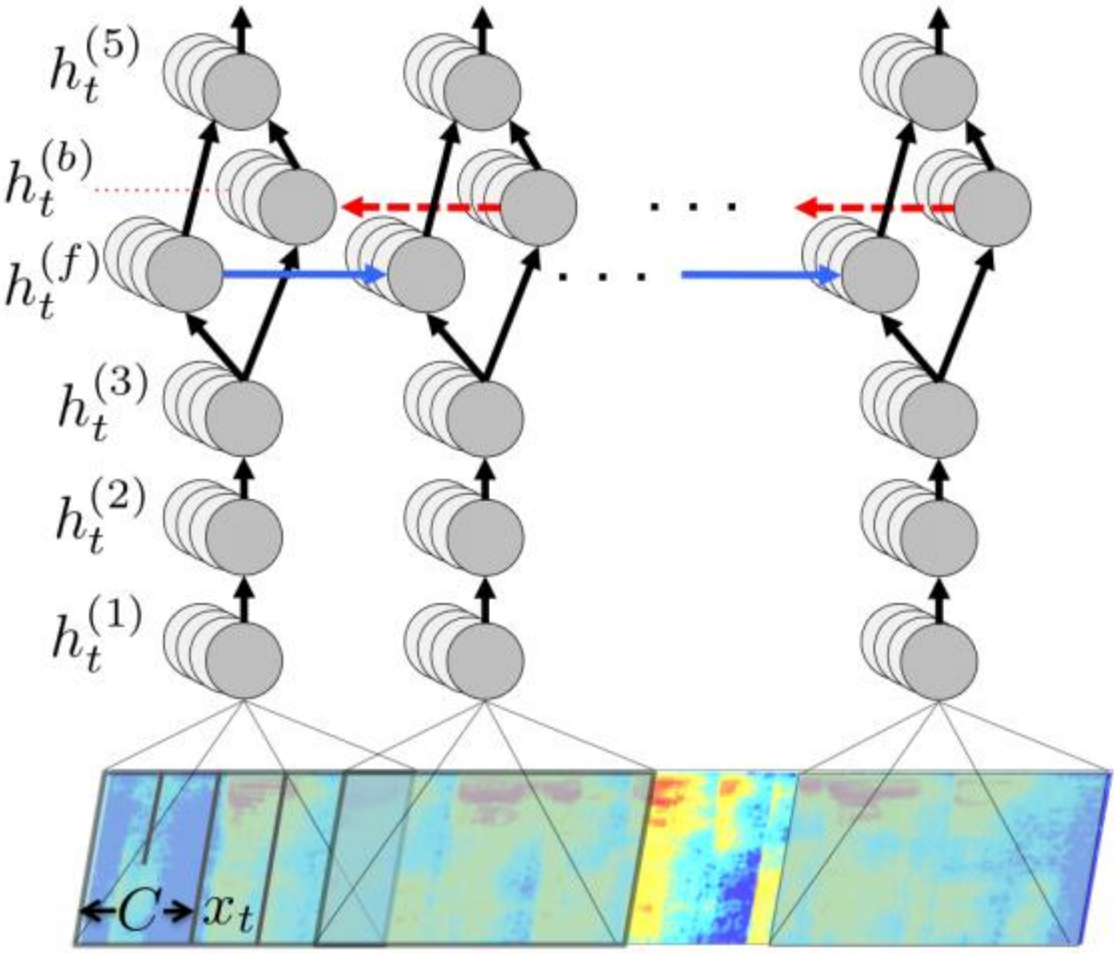
DL Meetup on July 9 2016

Deep Learning Models on Continuous/Dense Data



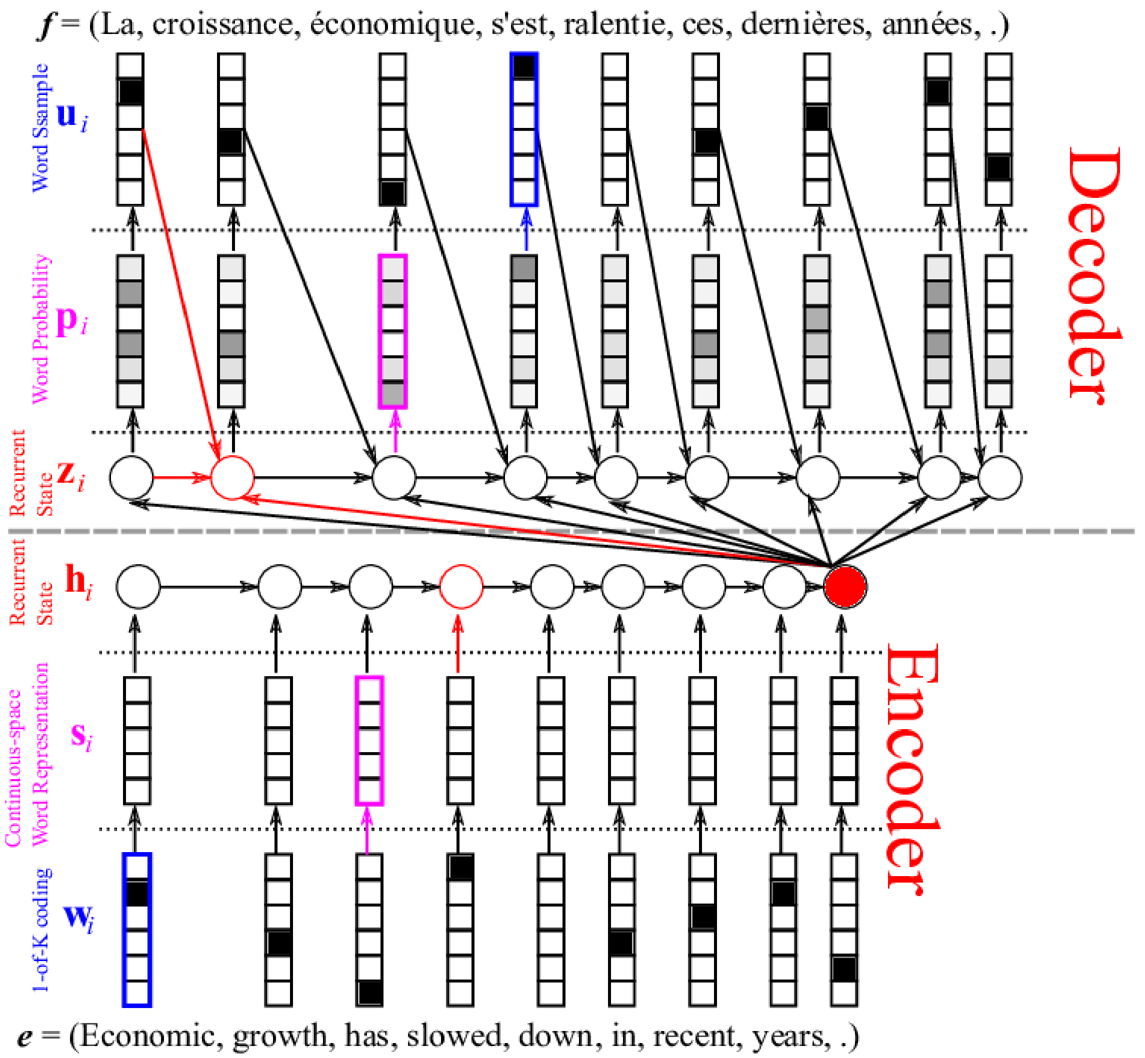
[Source: http://deeplearning.net/tutorial/lenet.html]

Deep Learning Models on Continuous/Dense Data



[Awni Hannun et al. Deep Speech: Scaling up end-to-end speech recognition. ArXiv 1412.5567 2014.]

Deep Learning Models on Continuous/Dense Data



[Kyunghyun Cho. Introduction to Neural Machine Translation with GPUs. 2015.]

Continuous dense data vs. Categorical Sparse Data

- Deep neural networks work well on continuous dense or sequence data
  - E.g., image pixels, audio, word sequences
  - Good to explore non-trivial local data patterns

- What if learning on multi-field categorical data?
  - E.g., [Weekday=Wednesday, Gender=Male, City=Shanghai,…]
  - Each field data has no explicit dependency on other fields

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

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 occupation: Student
- User tags: Sports, Electronics

➡ Click (1) or not (0)?

Predicted CTR (0.15)

Traditional One-Hot Binary Encoding

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

x=[0,0,1,0,0,0,0 0,1 0,0,1,0…0]

- High dimensional sparse binary feature vector
  - Usually higher than 1M dimensions, even 1B dimensions
  - Extremely sparse
- Impossible to directly deploy neural network models
  - E.g., input features 1M, first layer 500, then 500M parameters for first layer

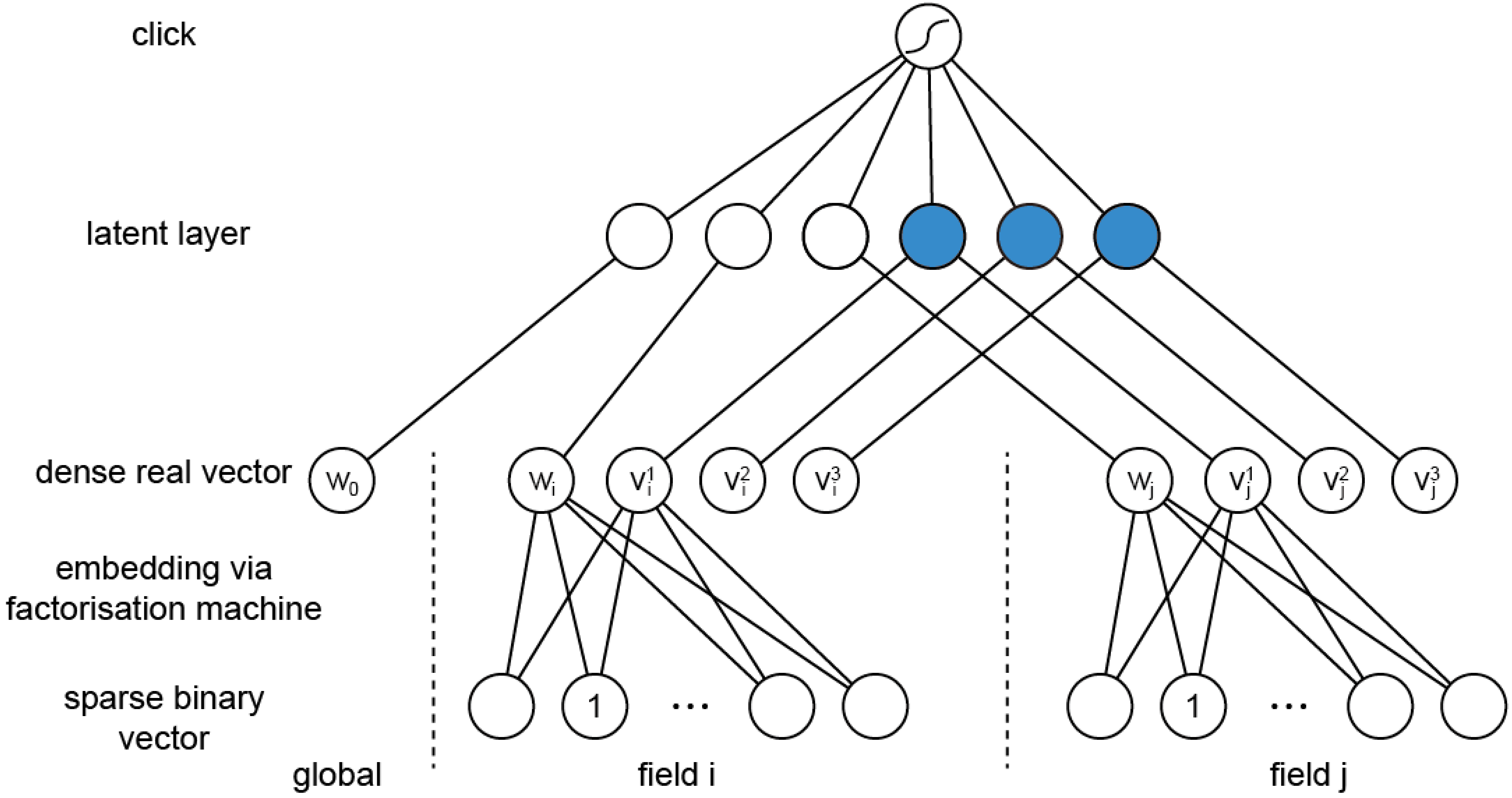
Leveraging Feature Embedding

- Factorisation Machines

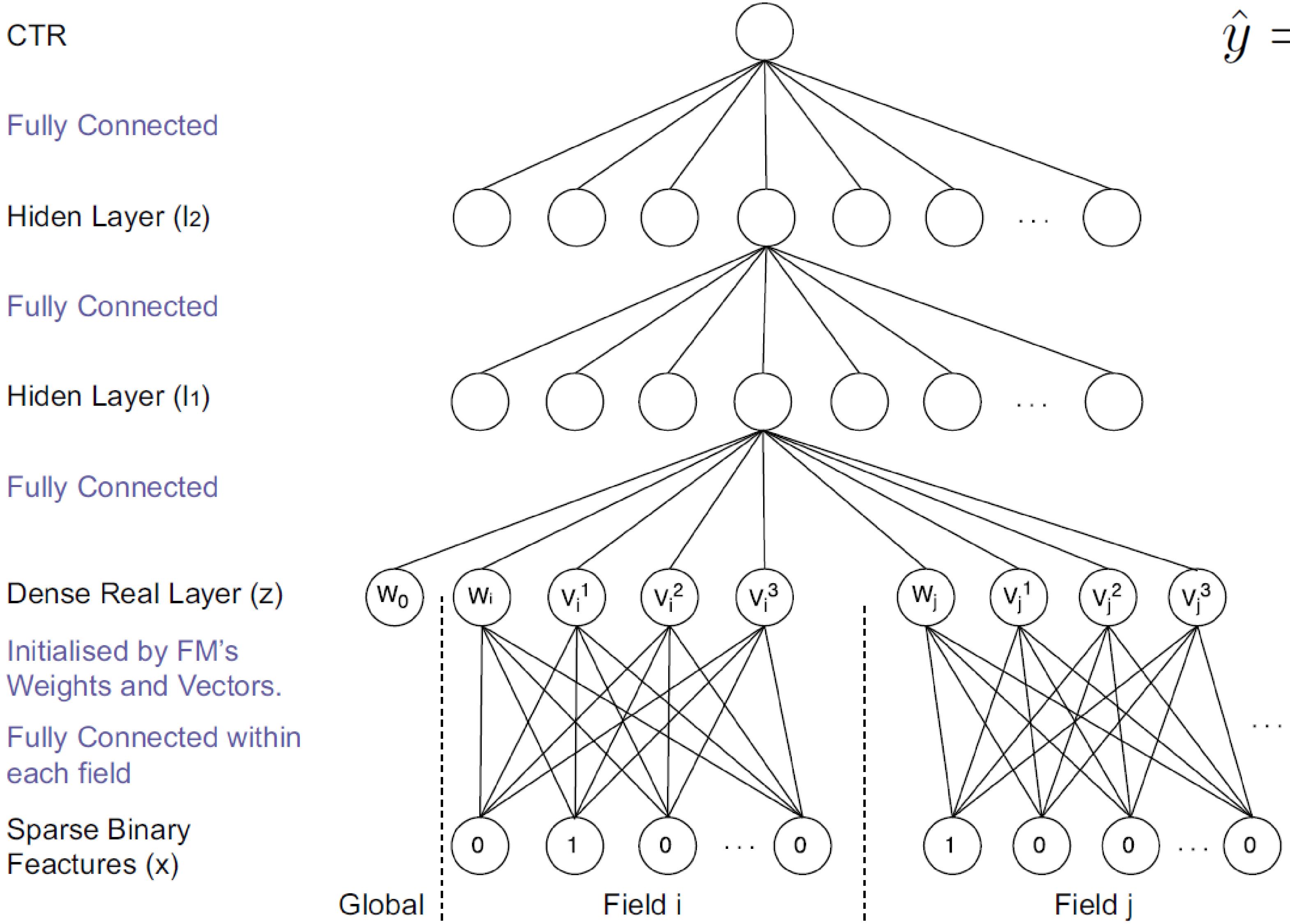
$$y_{\mathrm{FM}}(\boldsymbol{x}) := \mathrm{sigmoid}\Big(\underbrace{w_0 + \sum_{i=1}^{N} w_i x_i}_{\text{Logistic Regression}} + \underbrace{\sum_{i=1}^{N}\sum_{j=i+1}^{N} \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle x_i x_j}_{\text{Feature Interactions}}\Big)$$

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

Factorisation Machine is a Neural Network



Factorisation-machine supported Neural Networks (FNN)



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

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

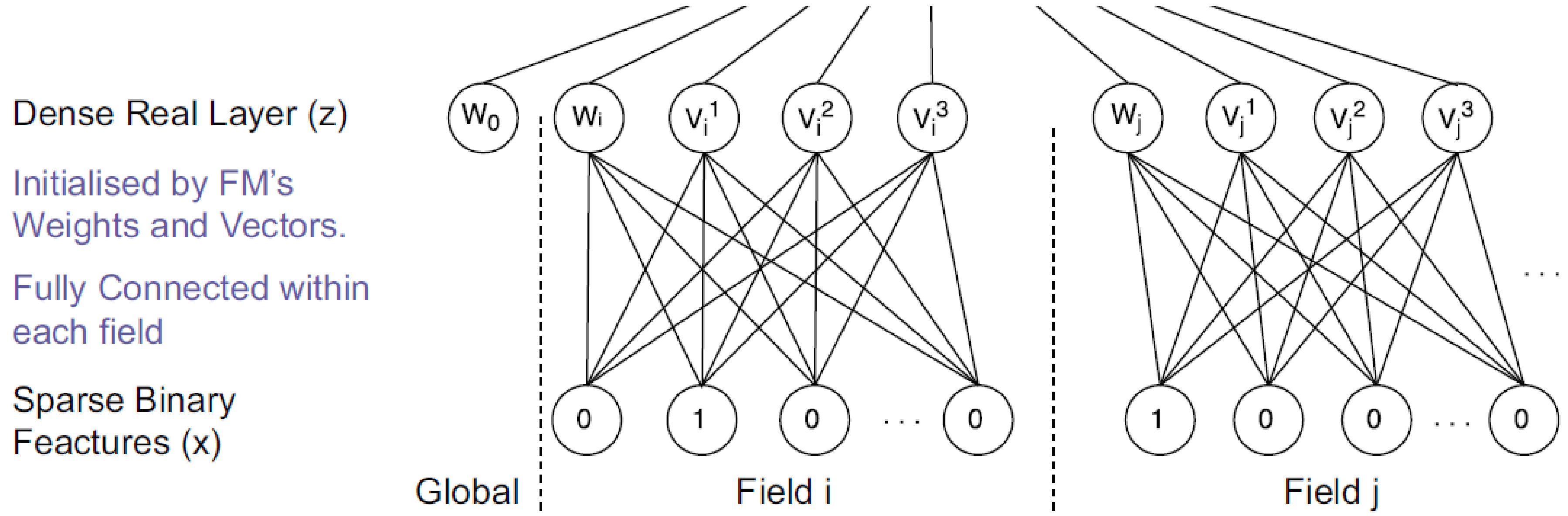
$$\boldsymbol{l}_1 = \tanh(\boldsymbol{W}_1 \boldsymbol{z} + \boldsymbol{b}_1)$$

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

[Factorisation Machine Initialised]

[Weinan Zhang et al. Deep Learning over Multi-Field Categorical Data: A Case Study on User Response Prediction. ECIR 2016]

Factorisation-machine supported Neural Networks (FNN)



- Chain rule to update factorisation machine parameters

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

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

[Weinan Zhang et al. Deep Learning over Multi-Field Categorical Data: A Case Study on User Response Prediction. ECIR 2016]

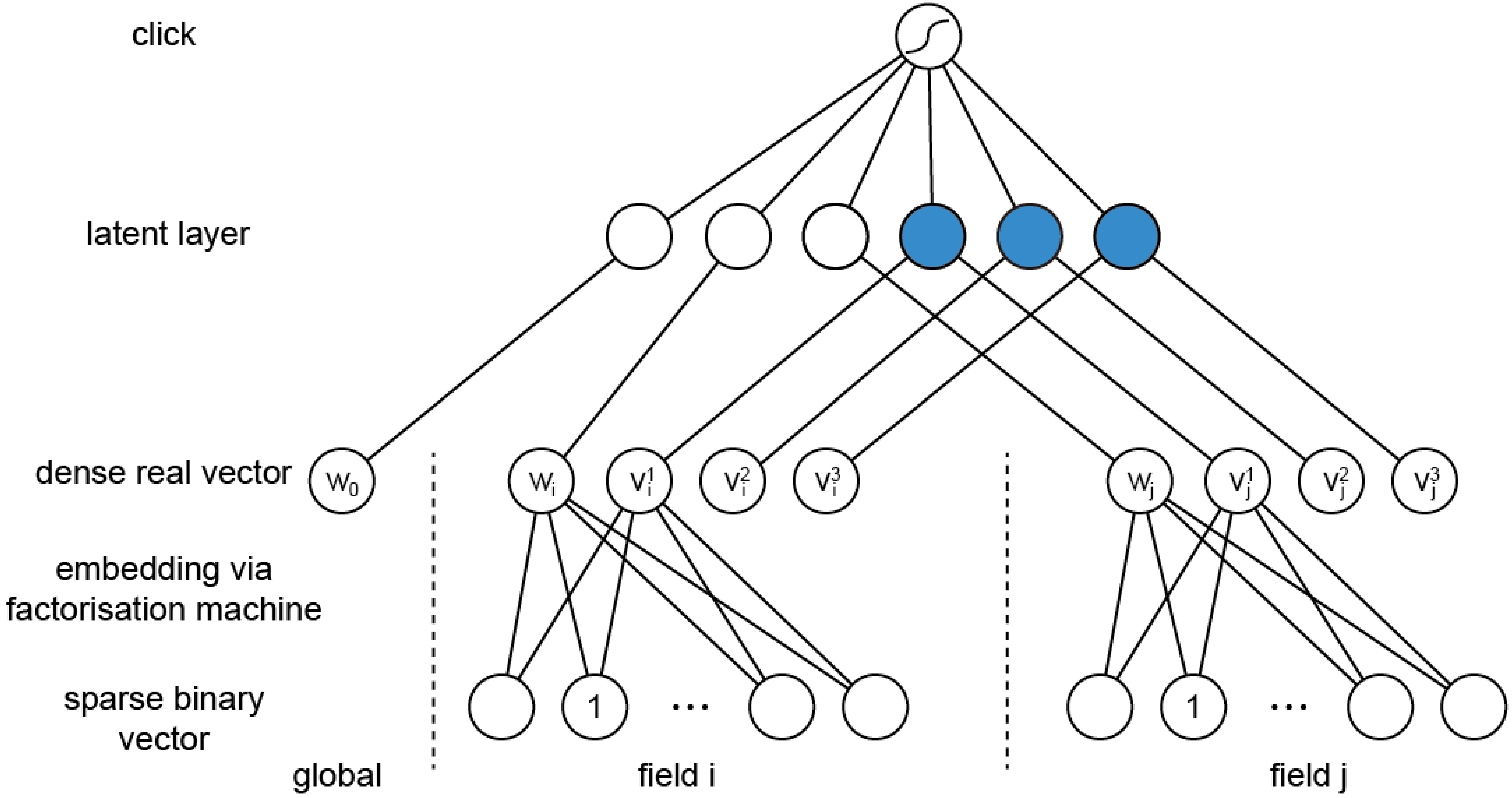
Overall CTR estimation AUC performance

- iPinYou dataset: 19.50M data instances

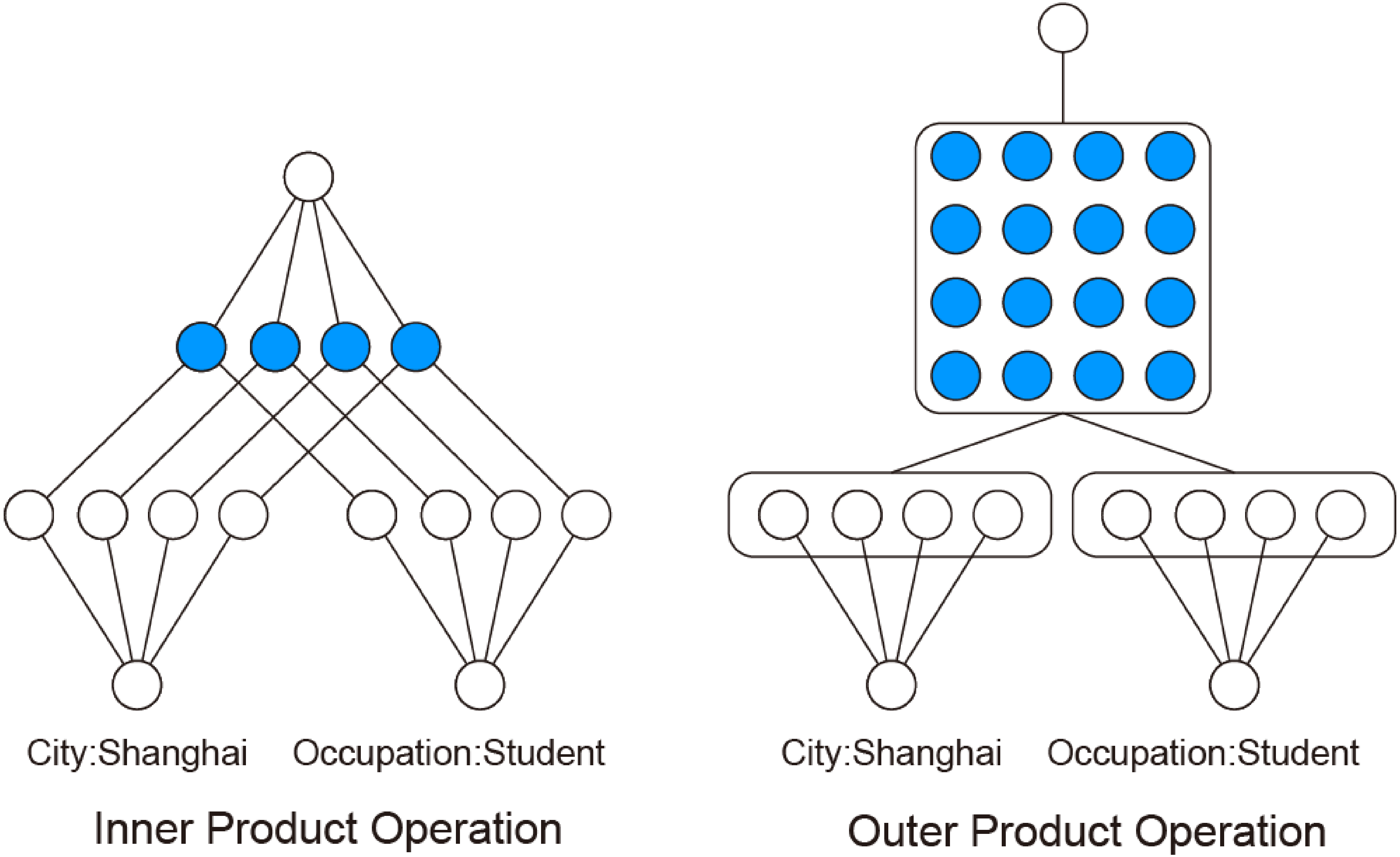
Campaign	LR	FM	FNN
1458	70.42%	70.21%	**70.52%**
2259	69.66%	69.73%	**69.74%**
2261	62.03%	60.97%	**62.99%**
2997	60.77%	60.87%	**61.41%**
3386	80.30%	79.05%	**80.56%**
all	68.81%	68.18%	**70.70%**

[Weinan Zhang et al. Deep Learning over Multi-Field Categorical Data: A Case Study on User Response Prediction. ECIR 2016]

But factorisation machine is still different from common additive neural networks

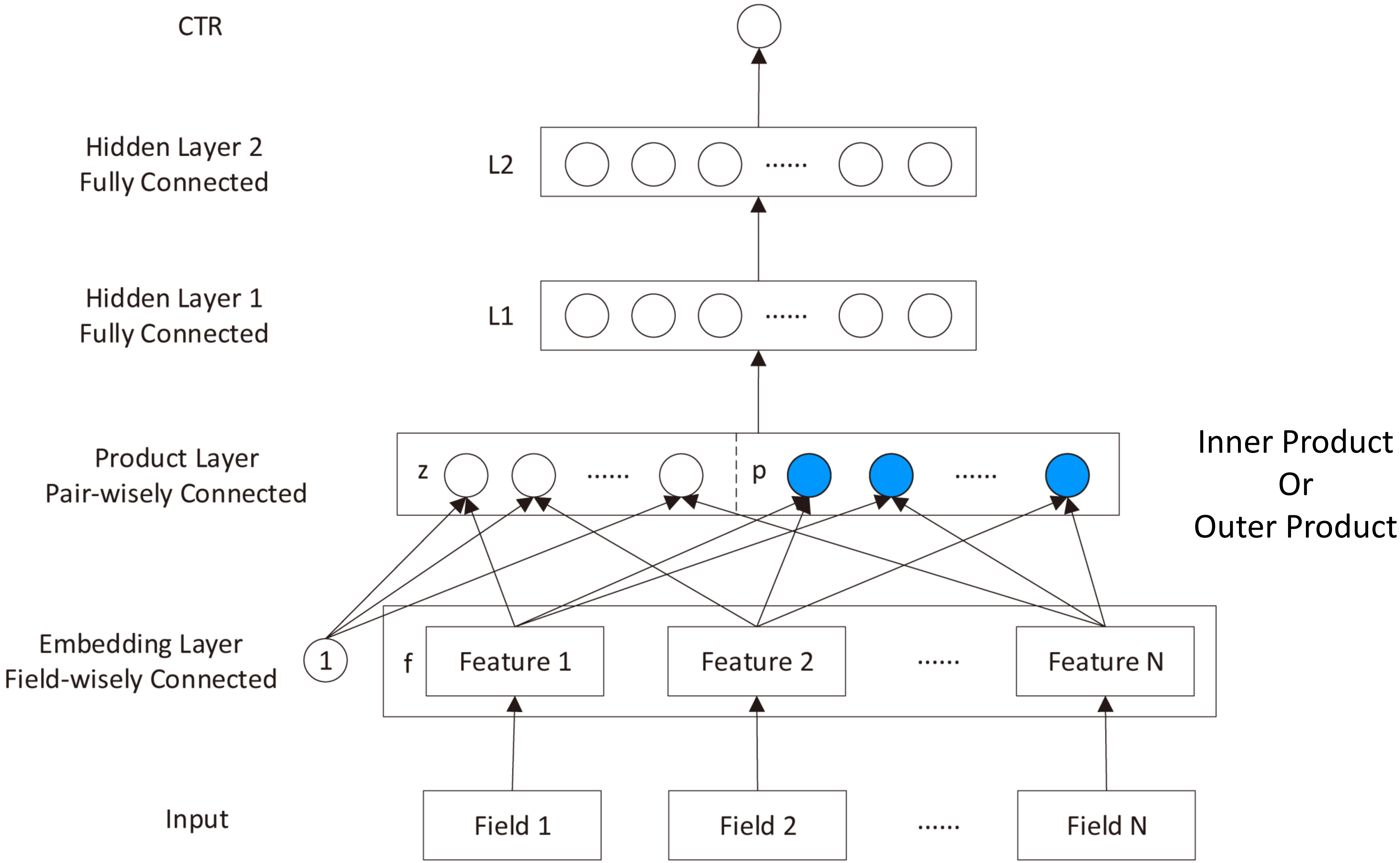


Product Operations as Feature Interactions



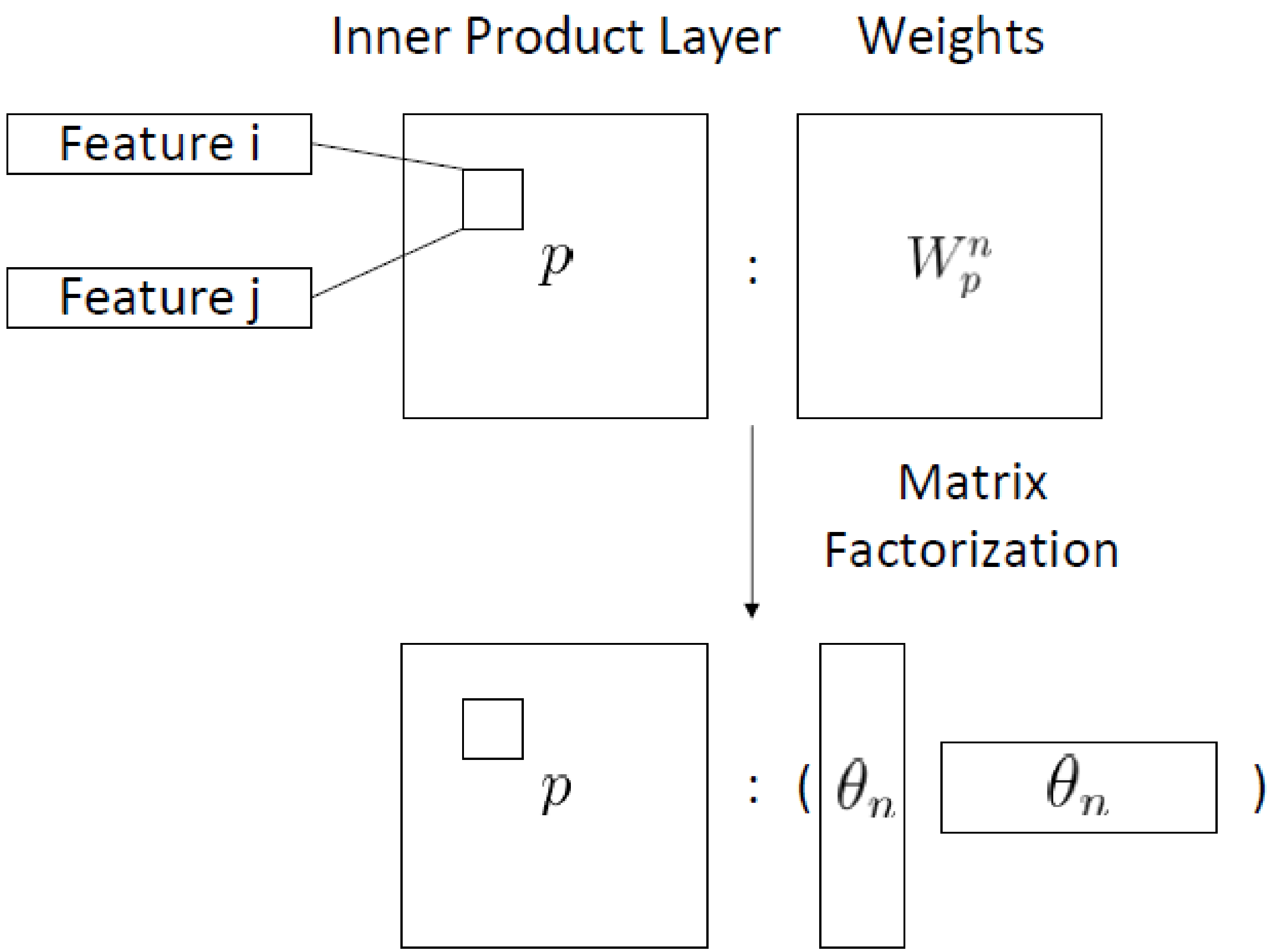
[Yanru Qu et al. Product-based Neural Networks for User Response Prediction. Paper in submission 2016]

PNN



[Yanru Qu et al. Product-based Neural Networks for User Response Prediction. Paper in submission 2016]

Some Tricks to Reduce the Complexity



[Yanru Qu et al. Product-based Neural Networks for User Response Prediction. Paper in submission 2016]

Experimental Results

Datasets

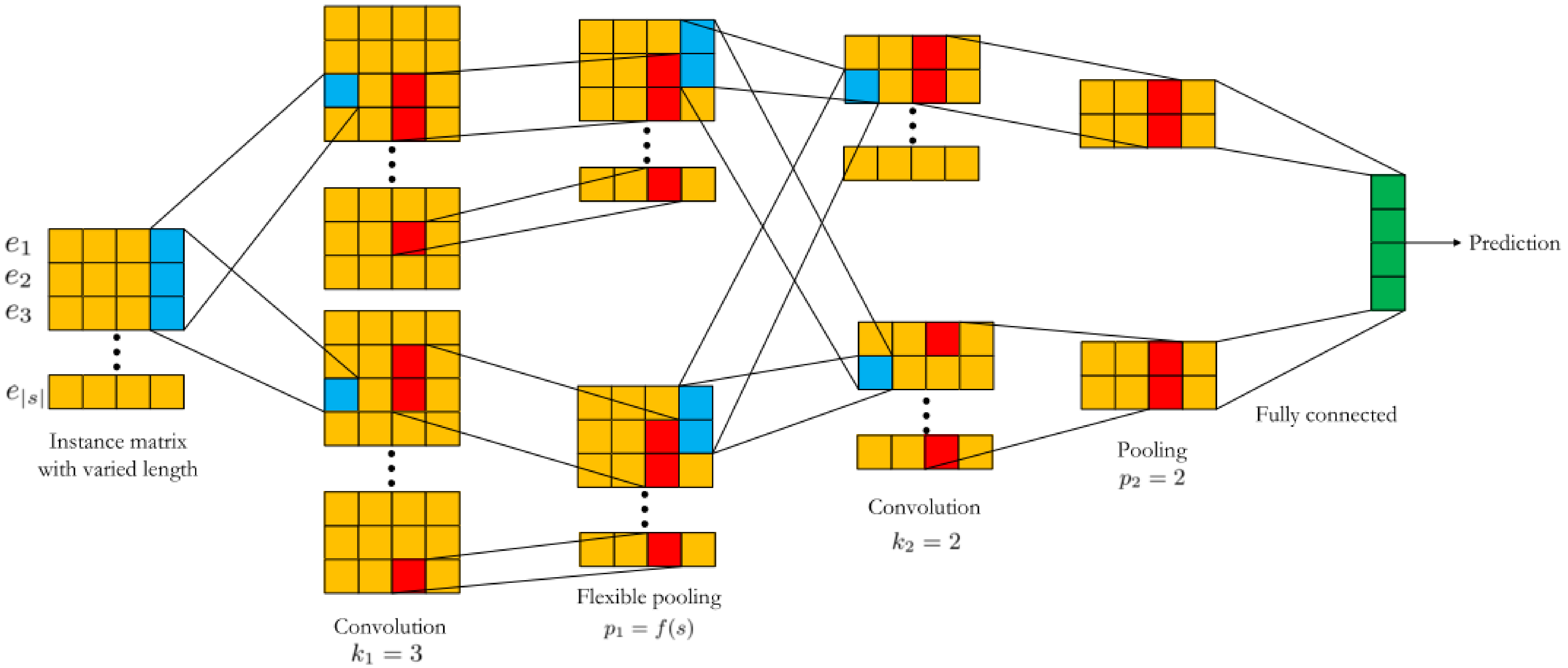
- Criteo Terabyte Dataset
  - 13 numerical fields, 26 categorical fields
  - 7 consecutive days out of 24 days in total (about 300 GB) during 2014
  - 79.4M impressions, 1.6M clicks after negative down sampling

- iPinYou Dataset
  - 65 categorical fields
  - 10 consecutive days during 2013
  - 19.5M impressions, 937.7K clicks without negative down sampling

Compared Algorithms

- LR – Logistic regression
- FM – Factorisation machine
- FNN – Factorisation machine supported neural network
- CCPM – Convolutional click prediction model

Convolutional Click Prediction Model (CCPM)



[Qiang Liu et al. A convolutional click prediction model. CIKM 2015]

Compared Algorithms

- LR – Logistic regression
- FM – Factorisation machine
- FNN – Factorisation machine supported neural network
- CCPM – Convolutional click prediction model

- PNN-I – Inner product neural network
- PNN-II – Outer product neural network
- PNN-III – Inner&outer product ensembled neural network

Evaluation Metrics

- Area under ROC curve (AUC)
- Log loss

$$\mathrm{LogLoss} = \frac{1}{|D|}\sum_{i=1}^{|D|}(-y_i \log \hat{y}_i - (1 - y_i)\log(1 - \hat{y}_i))$$

- Root mean squared error (RMSE)

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{|D|}(y_i - \hat{y}_i)^2}{|D|}}$$

- Relative Information Gain (RIG)

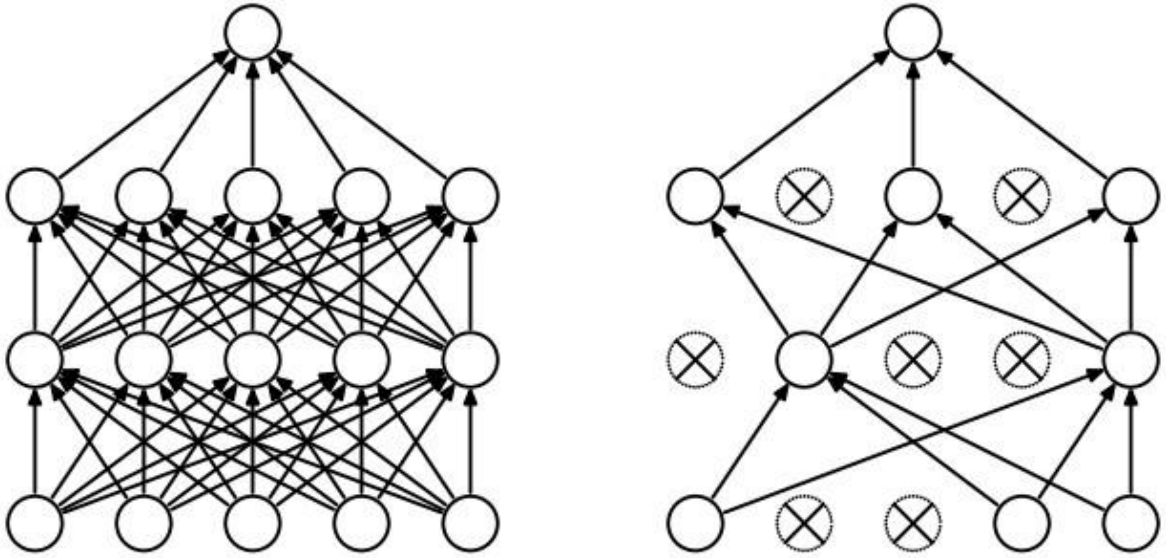
$$\mathrm{RIG} = 1 - \frac{\mathrm{LogLoss}}{-(p\log p + (1-p)\log(1-p))}$$

Criteo

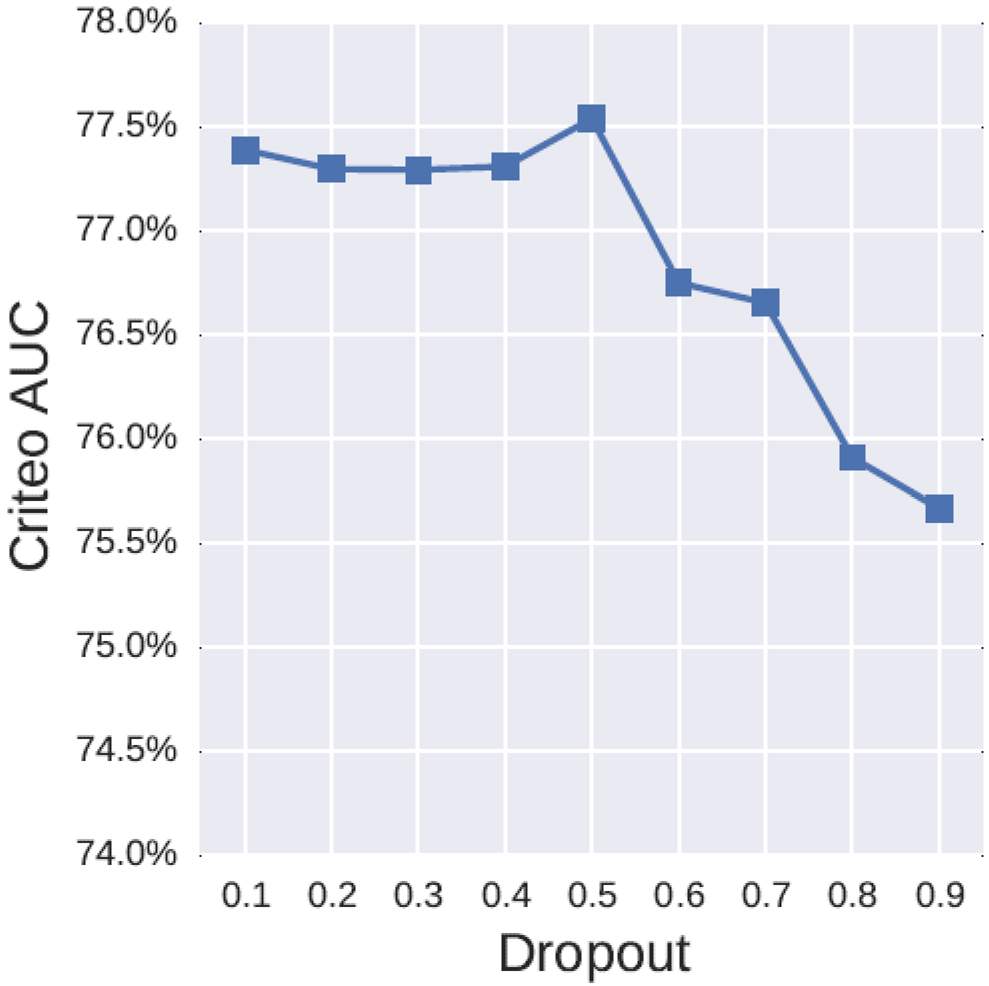
Model	AUC		Log Loss	
	Criteo	iPinYou	Criteo	iPinYou
LR	71.48%	73.43%	0.1334	5.581e-3
FM	72.20%	75.52%	0.1324	5.504e-3
FNN	75.66%	76.19%	0.1283	5.443e-3
CCPM	76.71%	76.38%	0.1269	5.522e-3
PNN-I	**77.79%**	79.14%	**0.1252**	5.195e-3
PNN-II	77.54%	**81.74%**	0.1257	5.211e-3
PNN-III	77.00%	76.61%	0.1270	**4.975e-3**

Model	RMSE		RIG	
	Criteo	iPinYou	Criteo	iPinYou
LR	9.362e-4	5.350e-07	6.680e-2	7.353e-2
FM	9.284e-4	5.343e-07	7.436e-2	8.635e-2
FNN	9.030e-4	5.285e-07	1.024e-1	9.635e-2
CCPM	8.938e-4	5.343e-07	1.124e-1	8.335e-2
PNN-I	**8.803e-4**	4.851e-07	**1.243e-1**	1.376e-1
PNN-II	8.846e-4	5.293e-07	1.211e-1	1.349e-1
PNN-III	8.988e-4	**4.819e-07**	1.118e-1	**1.740e-1**

[Yanru Qu et al. Product-based Neural Networks for User Response Prediction. Paper in submission 2016]

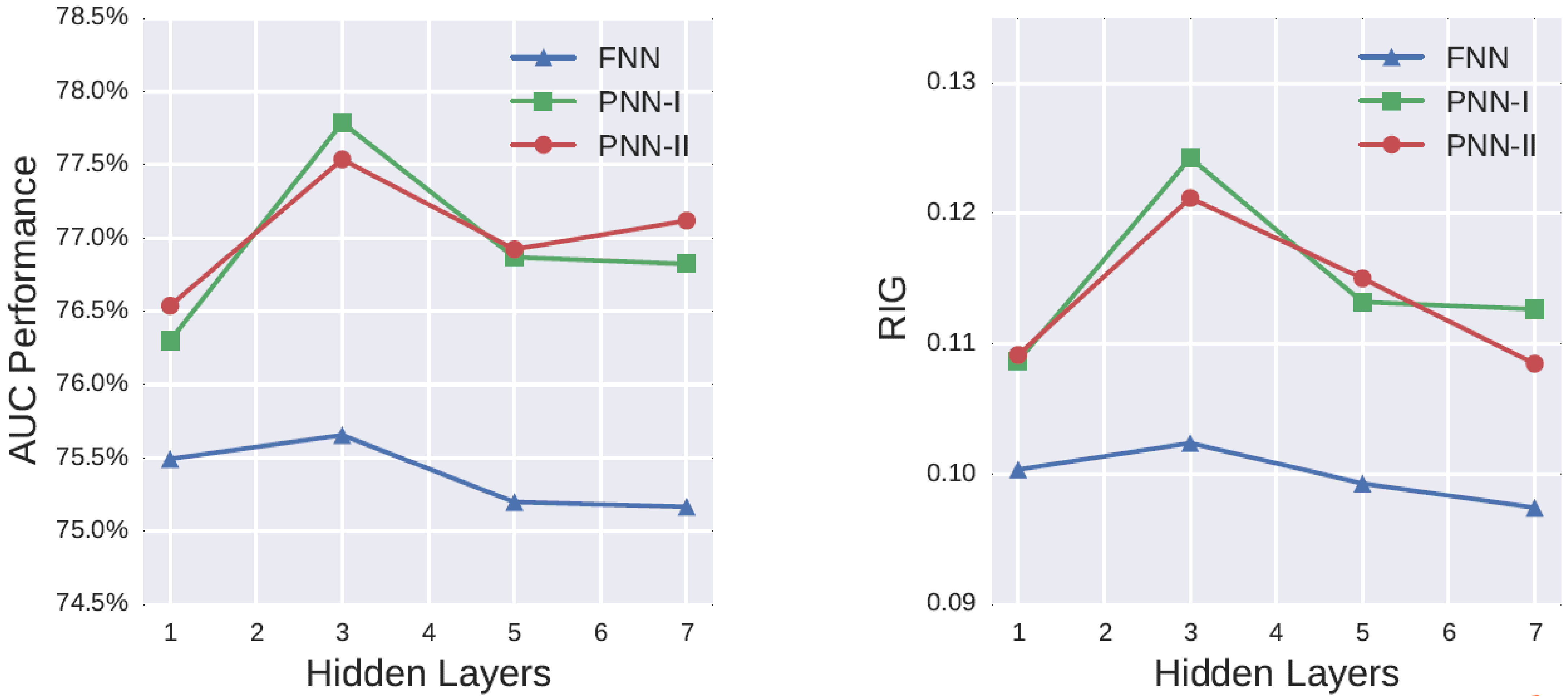


(a) Standard Neural Net (b) After applying dropout.



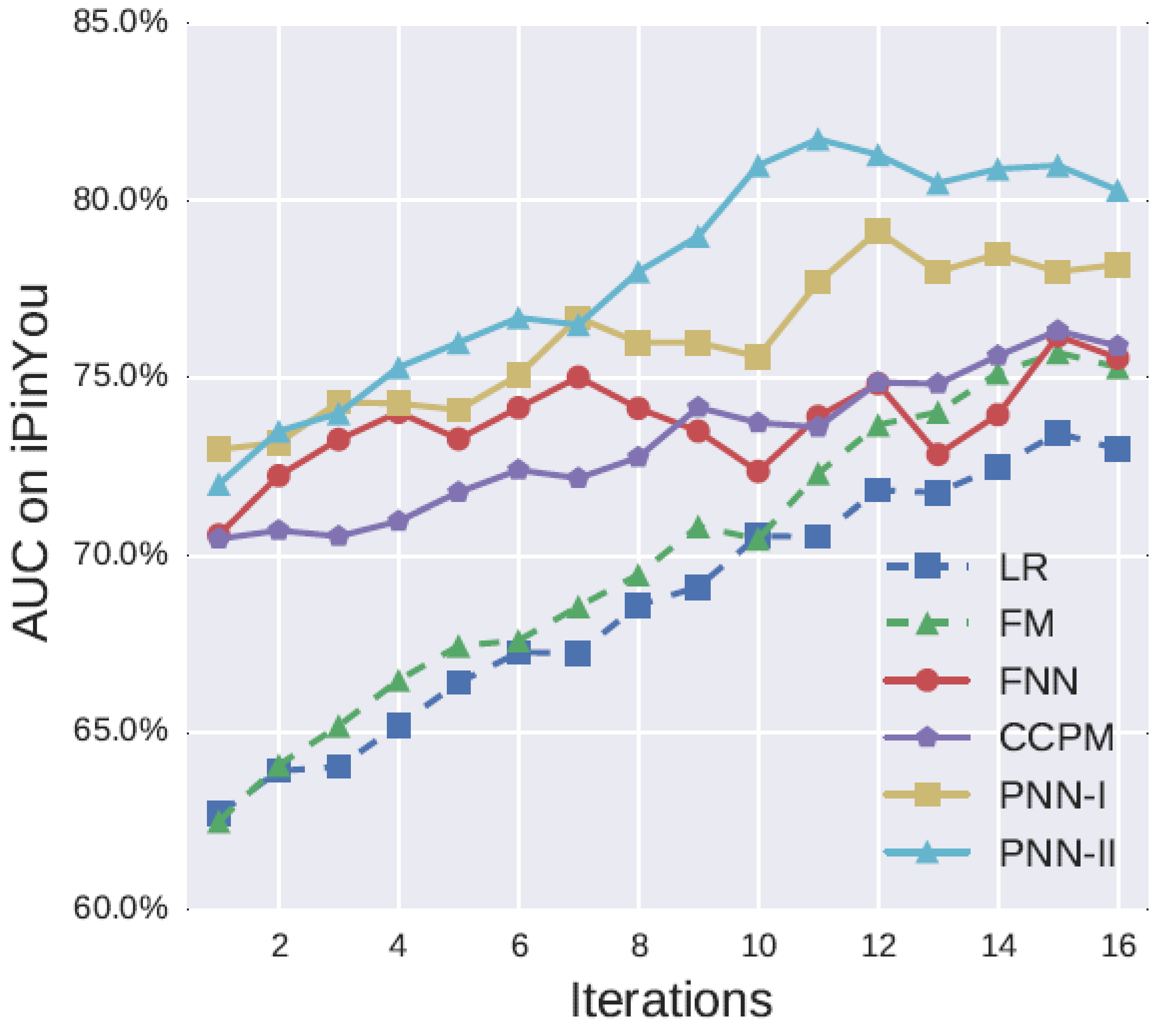
AUC Comparison of Dropout (PNN-II)

[Yanru Qu et al. Product-based Neural Networks for User Response Prediction. Paper in submission 2016]



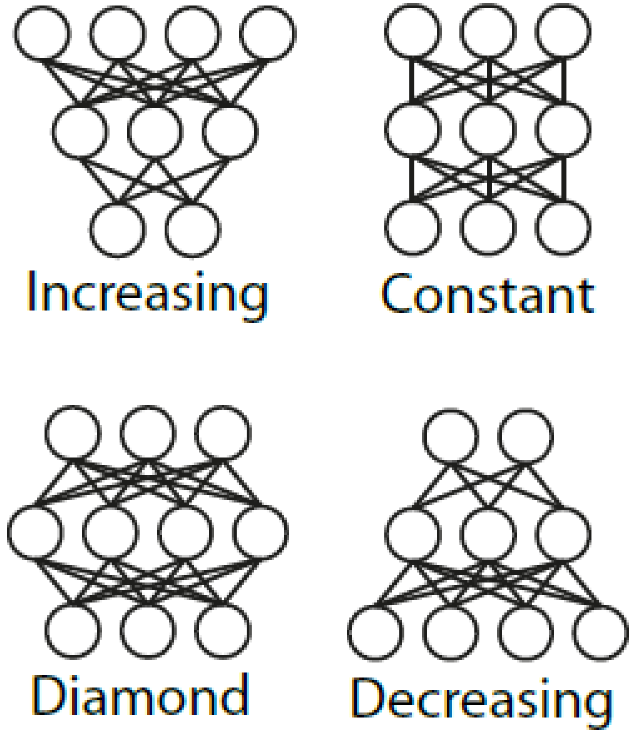
Performance comparison on network depth

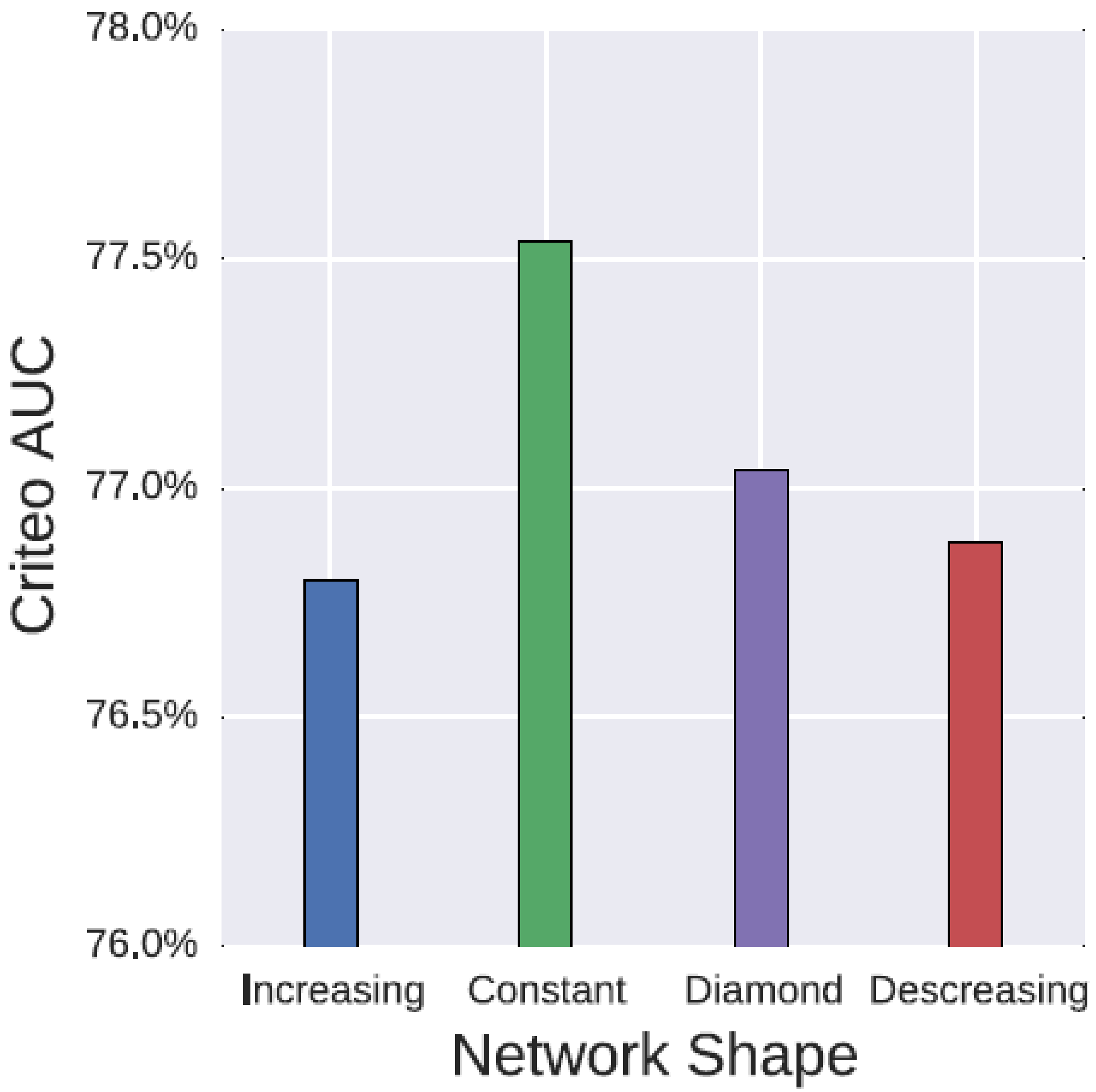
[Yanru Qu et al. Product-based Neural Networks for User Response Prediction. Paper in submission 2016]



Training Curve on iPinYou Dataset

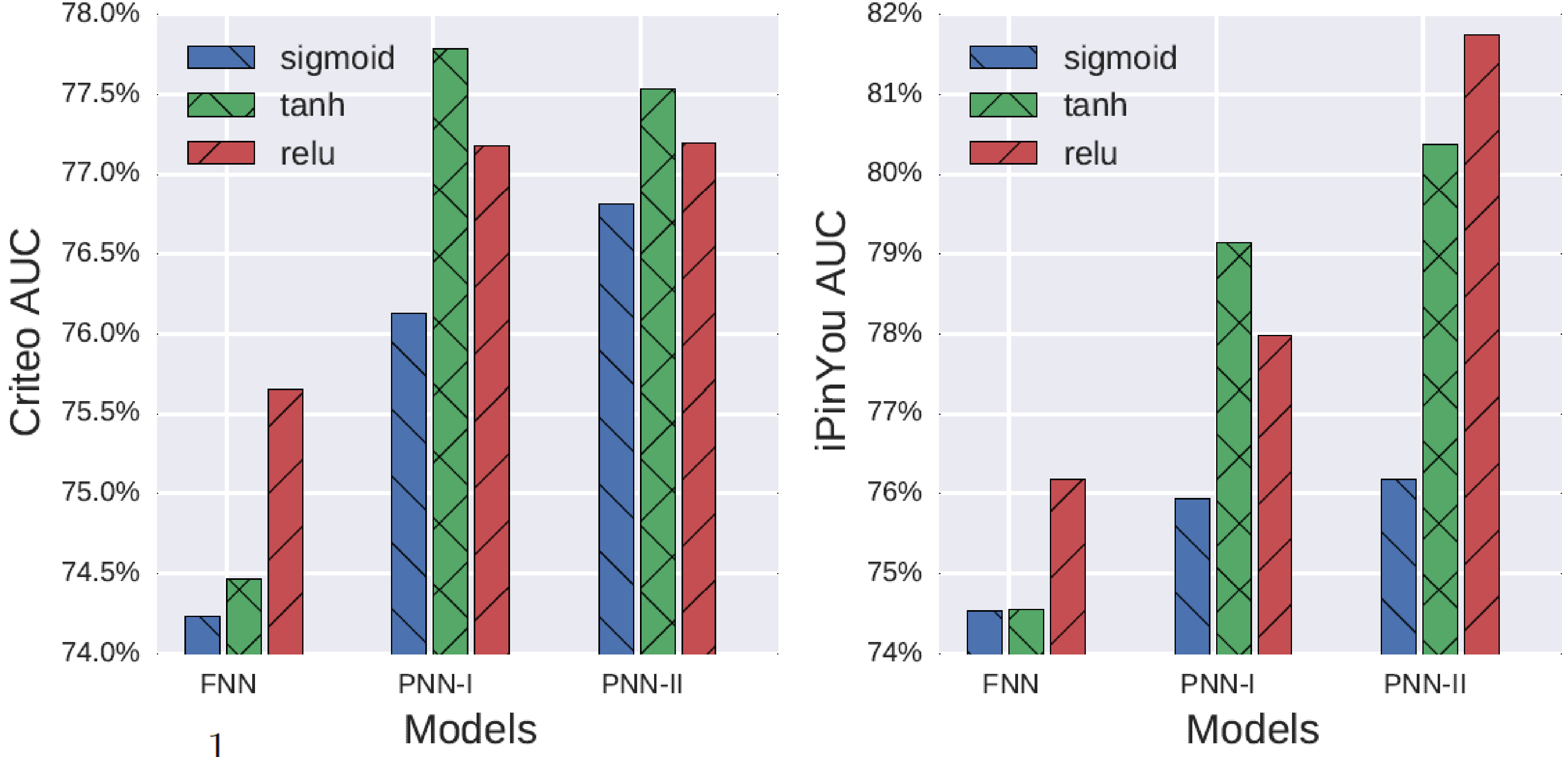
[Yanru Qu et al. Product-based Neural Networks for User Response Prediction. Paper in submission 2016]





AUC Comparison of Network Shape (PNN-II)

[Yanru Qu et al. Product-based Neural Networks for User Response Prediction. Paper in submission 2016]



$$\sigma(x) = \frac{1}{1+e^{-x}}$$

$$\tanh(x) = \frac{1-e^{-2x}}{1+e^{-2x}}$$

$$\mathrm{relu}(x) = \max(0, x)$$

AUC Comparison of Activation Functions

[Yanru Qu et al. Product-based Neural Networks for User Response Prediction. Paper in submission 2016]

Summary

- An attempt of exploring deep neural network efficacy on multi-field categorical data

- Layers of product operations are introduced to model the inter-field feature interactions

- On display ad CTR estimation tasks, PNNs significantly outperform other models

A New Deep Learning Paradigm: Deep learning over ID format data

Neu-IR: The SIGIR 2016 Workshop on Neural Information Retrieval

The first international Neu-IR (pronounced *"new IR"*) workshop on neural information retrieval will be hosted at SIGIR 2016 in Pisa, Tuscany, Italy on 21st July, 2016.

To get the latest *updates* or ask any *questions* please follow us on twitter.

neural
information
retrieval
2016

Nick Craswell
Microsoft
Bellevue, US

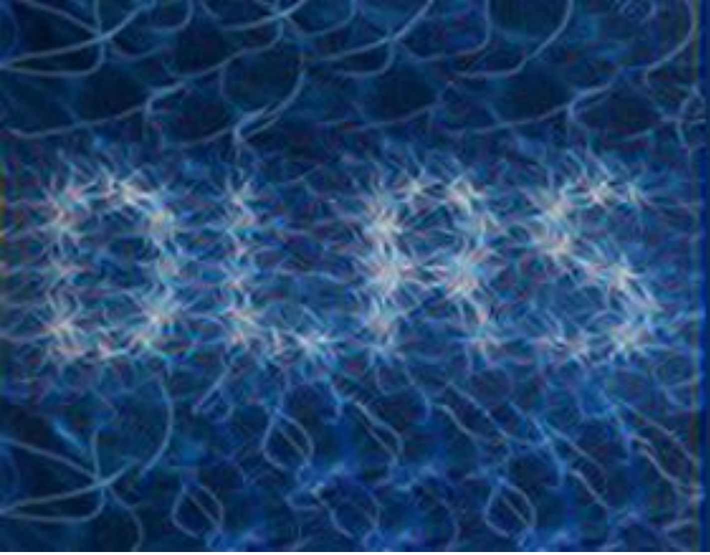
W. Bruce Croft
University of Massachusetts
Amherst, US

Maarten de Rijke
University of Amsterdam
Amsterdam, The Netherlands

Jiafeng Guo
Chinese Academy of Sciences
Beijing, China

Bhaskar Mitra
Microsoft
Cambridge, UK

[https://www.microsoft.com/en-us/research/event/neuir2016/]



1st Workshop on Deep Learning for Recommender Systems

in conjunction with RecSys 2016
15 September 2016, Boston, USA

Goals & Motivation

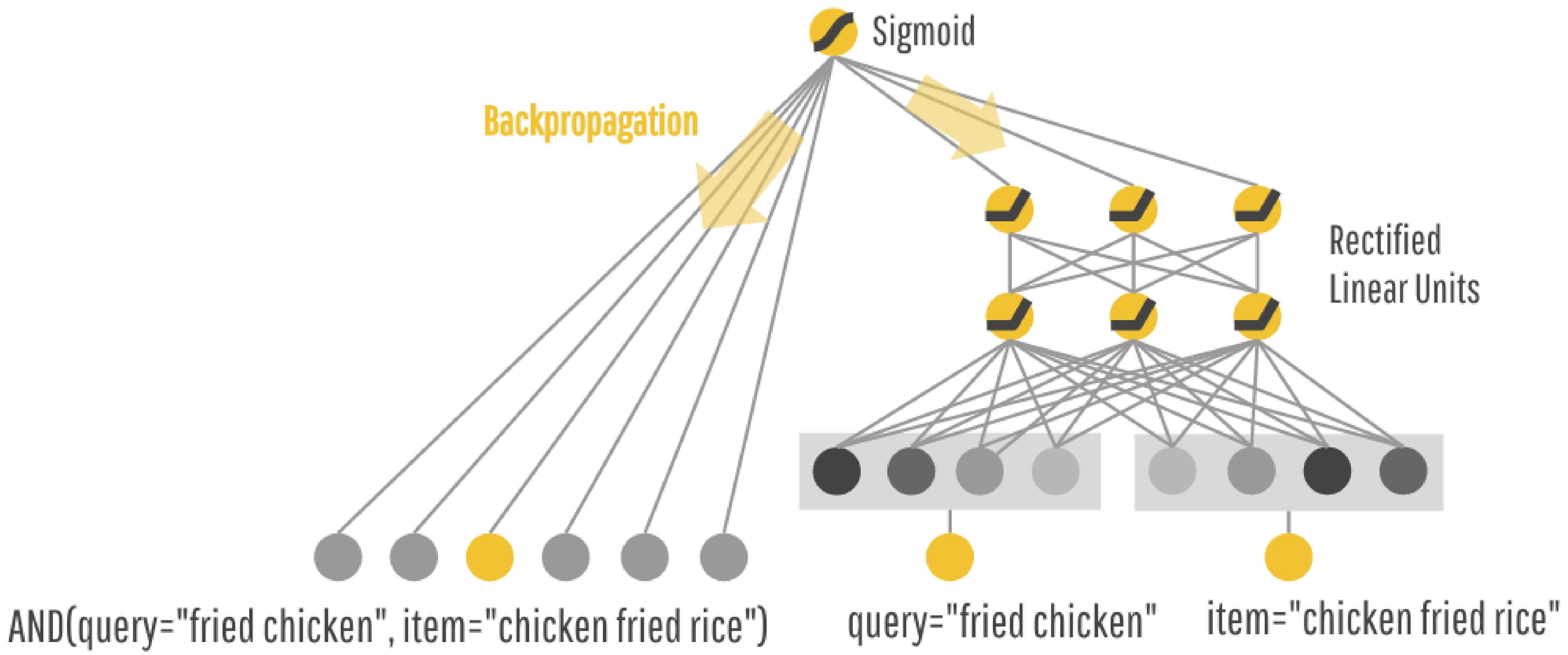
We believe that Deep Learning is one of the next big things in Recommendation Systems technology. The past few years have seen the tremendous success of deep neural networks in a number of complex tasks such as computer vision, natural language processing and speech recognition. Despite this, only little work has been published on Deep Learning methods for Recommender Systems. Notable recent application areas are music recommendation, news recommendation, and session-based recommendation. The aim of the workshop is to encourage the application of Deep Learning techniques in Recommender Systems, to promote research in deep learning methods for Recommender Systems, and to bring together researchers from the Recommender Systems and Deep Learning communities.

ORGANIZERS

- Alexandros Karatzoglou, Telefonica Research, Spain
- Balázs Hidasi, Gravity R&D, Hungary
- Domonkos Tikk, Gravity R&D, Hungary
- Oren Sar-Shalom, IBM Research, Israel
- Haggai Roitman, IBM Research, Isreal
- Bracha Shapira, Ben-Gurion University, Isreal
- Lior Rokach, Ben-Gurion University, Isreal

[http://dlrs-workshop.org/]

Wide and Deep Learning by Google



[Heng-Tze Cheng et al. Wide & Deep Learning for Recommender Systems. ArXiv 2016.]

Thank You. Questions?

- Key references
  - Weinan Zhang et al. Deep Learning over Multi-Field Categorical Data: A Case Study on User Response Prediction. ECIR 2016.
  - Qiang Liu et al. A convolutional click prediction model. CIKM 2015.
  - Yanru Qu et al. Product-based Neural Networks for User Response Prediction. Paper in submission 2016.
  - Heng-Tze Cheng et al. Wide & Deep Learning for Recommender Systems. ArXiv 2016.