CFM: Convolutional Factorization Machines for Context-Aware Recommendation

[1]Xin Xin, [2]Bo Chen, [3]Xiangnan He, et al.

[1]School of Computing Science, University of Glasgow
[2]School of Software Engineering, Shanghai Jiao tong University
[3]School of Data Science, University of Science and Technology of China

Presented by **Xin Xin**@IJCAI19, Aug.16, 2019

Factorization Machines (FM)

- FM [Rendel et al., ICDM2010] is one of the most effective feature-based recommendation algorithms

Factorization Machines (FM)

- FM [Rendel et al., ICDM2010] is one of the most effective feature-based recommendation algorithms
- One/Multi-hot feature vectors as inputs
  - Encodes both item/user side information and context information

$$\underbrace{[0,0,0,1,0,0,0]}_{\text{weekday=Thursday}} \quad \underbrace{[0,1,...,0]}_{\text{location=London}} \quad \underbrace{[1,1,0,...,0]}_{\text{historical items (multi-hot)}}$$

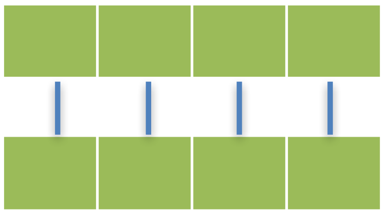
Factorization Machines (FM)

- FM [Rendel et al., ICDM2010] is one of the most effective feature-based recommendation algorithms
- One/Multi-hot feature vectors as inputs
- Combines linear regression and second-order feature interaction

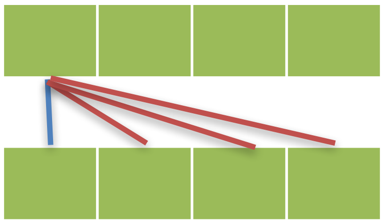
$$\hat{y}_{FM}(\mathbf{x}) = \underbrace{w_0 + \sum_{i=1}^{m} w_i \overset{\text{feature}}{\underline{x_i}}}_{\text{linear regression}} + \underbrace{\sum_{i=1}^{m}\sum_{j=i+1}^{m} x_i x_j \cdot \langle \overset{\text{embedding for } x_i}{\underline{\mathbf{v}_i}}, \mathbf{v}_j \rangle}_{\text{second-order feature interaction}}$$

Limitations of FM

- Inner product based feature interaction
  - Embedding dimensions are independent with each other



Inner product



?

  - There may be correlations between different dimensions [Zhang et al., SIGIR2014]
- Higher-order interaction & Non-linearity
  - NFM [He et al., SIGIR2017]
  - DeepFM [Guo et al., IJCAI2017]

Contributions

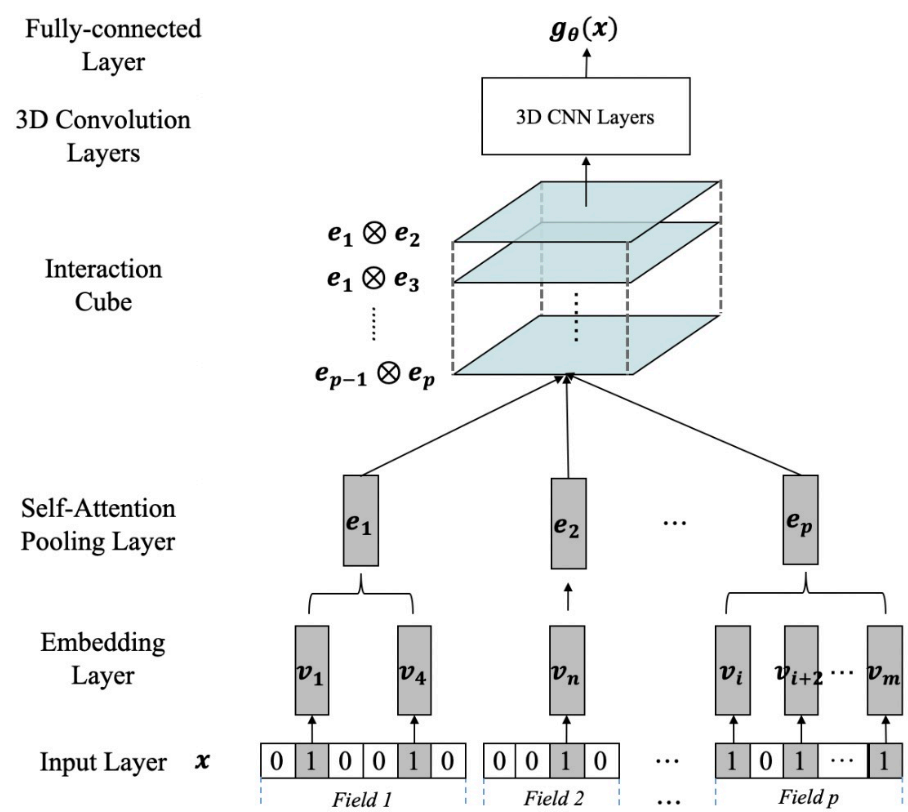
- Utilize an outer product-based interaction cube to represent feature interactions, which encodes both interaction signals and dimension correlations.
- Employ 3D CNN above the interaction cube to capture high-order interactions in an explicit way.
- Leverage an attention mechanism to perform feature pooling, reducing time complexity.

Convolutional Factorization Machines (CFM)

- Prediction rule:

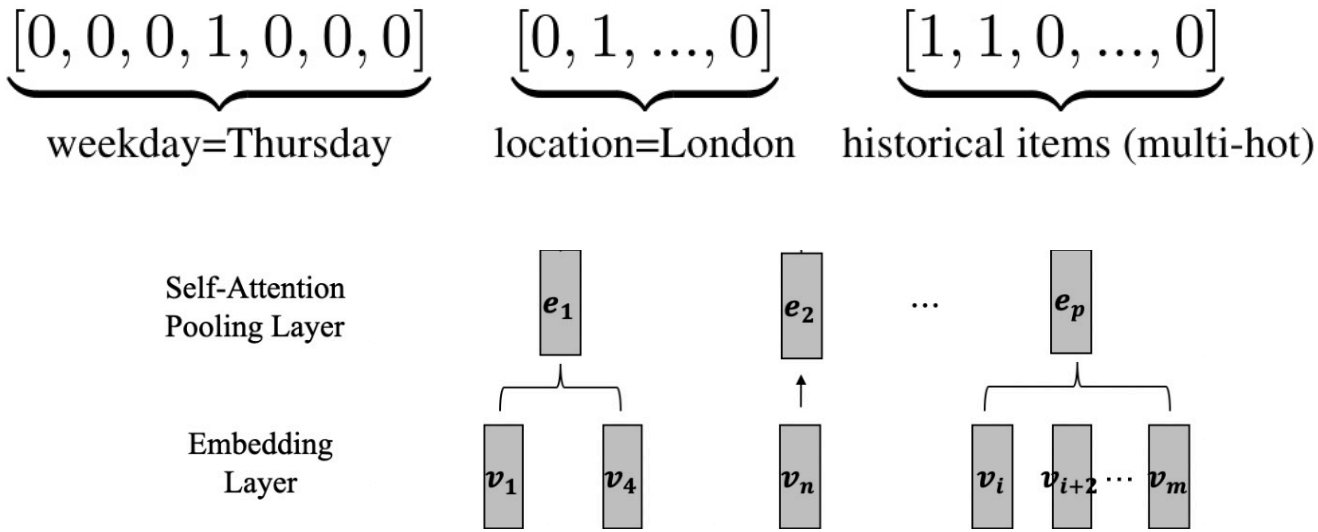
$$\hat{y}_{CFM}(\mathbf{x}) = w_0 + \sum_{i=1}^{m} w_i x_i + g_\theta(\mathbf{x})$$

- Overall structure:



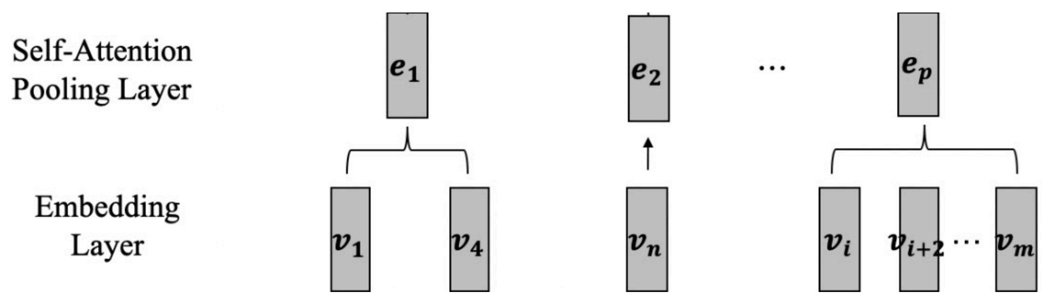
CFM

- Input and Embedding Layer
  - sparse feature vectors==>embedding table lookup
- Attention pooling layer



CFM

- Input and Embedding Layer
  - sparse feature vectors==>embedding table lookup
- Attention pooling layer

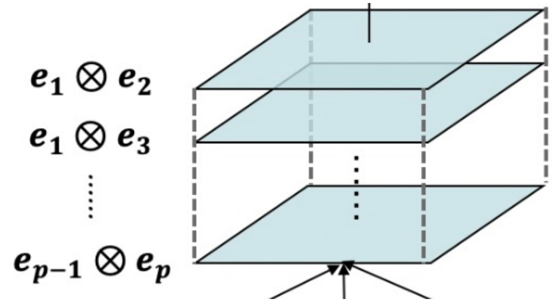


  - attention score $a_i = \mathbf{h}_j^T tanh(\mathbf{W}_j \mathbf{v}_i + \mathbf{b}_j)$,
  - softmax $\alpha_i = softmax(a_i) = \frac{exp(a_i)}{\sum_{x_{i'} \in \mathcal{X}_j} exp(a_{i'})}$
  - weighted sum $\mathbf{e}_j = \sum_{x_i \in \mathcal{X}_j} \alpha_i \mathbf{v}_i$

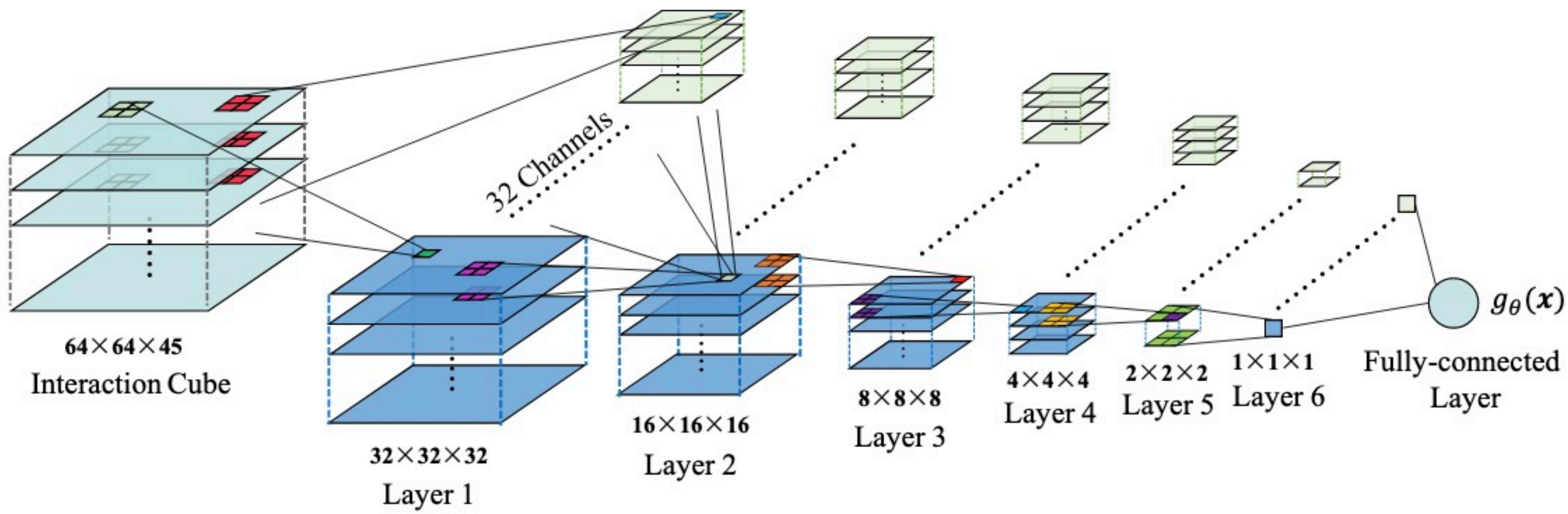
CFM

- Interaction Cube

$$\mathbf{M}_{i,j} = \mathbf{e}_i \otimes \mathbf{e}_j = \begin{bmatrix} e_{i1}e_{j1} & e_{i1}e_{j2} & \cdots & e_{i1}e_{jd} \\ e_{i2}e_{j1} & e_{i2}e_{j2} & \cdots & e_{i2}e_{jd} \\ \vdots & \vdots & \vdots & \vdots \\ e_{id}e_{j1} & e_{id}e_{j2} & \cdots & e_{id}e_{jd} \end{bmatrix}$$



- 3D CNN



CFM

- Model Training
  - Pair-wise ranking loss (BPR) [Rendle et al., UAI2009]

$$L = \sum -\ln \sigma(\hat{y}_{CFM}(\mathbf{x}^{+}) - \hat{y}_{CFM}(\mathbf{x}^{-})).$$

  - L2 regularization
  - Drop-out

Experiments

- Research questions:
  - Does CFM model outperform state-of-the-art methods for top-k recommendation?
  - How do the special designs of CFM (i.e., interaction cube and 3DCNN) affect the model performance?
  - What's the effect of the attention-based feature pooling?
- Datasets:
  - Frappe
  - Last.fm
  - MovieLens
- Evaluation:
  - Leave-one-out
  - HR&NDCG

Dataset	#users	#items	#transactions	#fields
Frappe	957	4,082	96,203	10
Last.fm	1,000	20,301	214,574	4
MovieLens	6,040	3,665	939,809	4

Experiments

- Baselines:
  - PopRank: popularity-based recommendation
  - FM[Rendle et al., ICDM2010]: original FM with BPR loss
  - NFM[He et al., SIGIR17]: stacking MLP upon FM
  - DeepFM[Guo et al., IJCAI2017]: wide&deep+FM
  - ONCF[He et al., IJCAI2018]: outer product+MF

Experiments

- RQ1 (performance)

Frappe	PopRank	FM	DeepFM	NFM	ONCF	CFM
HR@10	0.3493	0.5486	0.6035	0.6197	0.6531	**0.6720**
NDCG@10	0.1898	0.3469	0.3765	0.3924	0.4320	**0.4560**

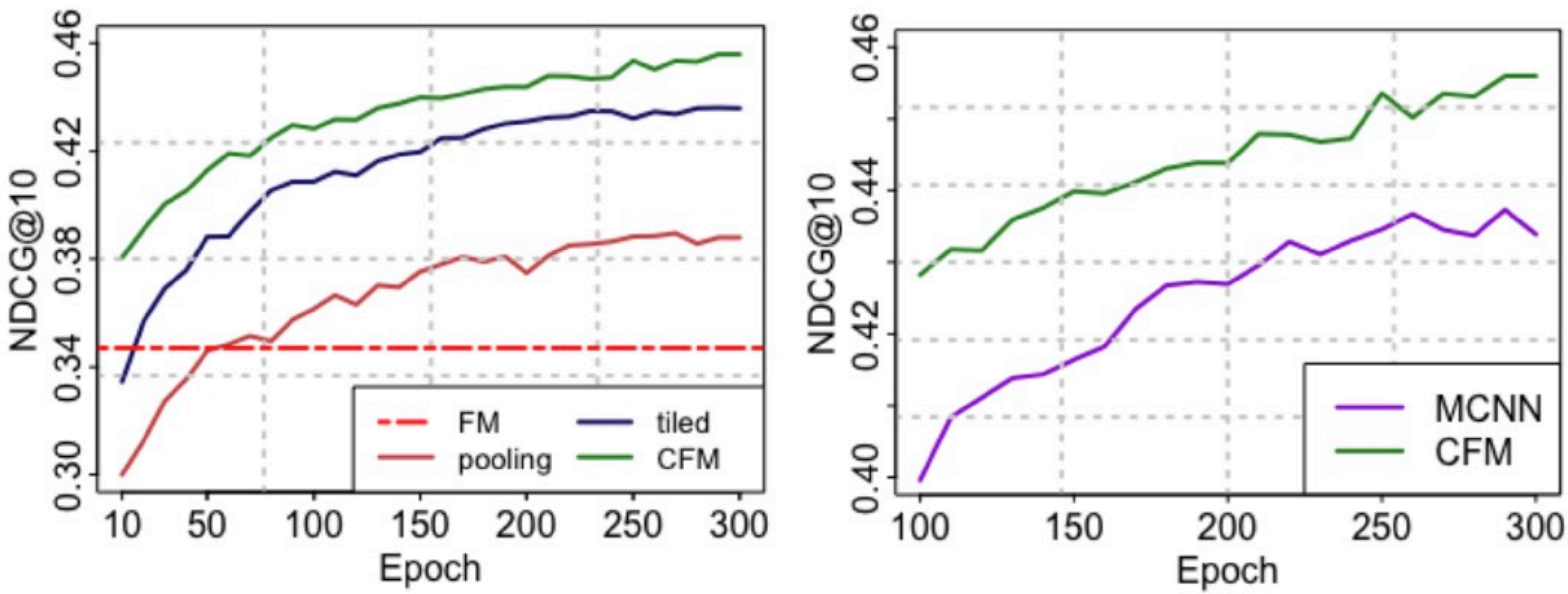
Last.fm	PopRank	FM	DeepFM	NFM	ONCF	CFM
HR@10	0.0023	0.2382	0.2612	0.2676	0.3208	**0.3538**
NDCG@10	0.0011	0.1374	0.1473	0.1488	0.1823	**0.1948**

Frappe	PopRank	FM	DeepFM	NFM	ONCF	CFM
HR@10	0.0235	0.0998	0.1170	0.1192	0.1110	**0.1323**
NDCG@10	0.0107	0.0452	0.0526	0.0553	0.0514	**0.0627**

  - Deep structure helps to improve FM (DeepFM&NFM)
  - CFM achieves the best performance

Results

- RQ2(model ablation)
  - Interaction cube & 3D CNN



(a) interaction cube (b) 3D CNN

3D architecture helps to improve performance

Results

- RQ3(feature pooling)
  - Effect of attention

Model	HR@10	HR@20	NG@10	NG@20
Max	0.1257	0.2142	0.0591	0.0813
Mean	0.1291	0.2212	0.0603	0.0833
CFM	0.1323	0.2248	0.0627	0.0858

  - Run time & performance

Indicator	FM	CFM	CFM-wfp
Time(min)	0.53	4.63	50.52
HR@10	0.0998	0.1323	0.1297
NG@10	0.0452	0.0627	0.0605

Attention pooing layer helps to improve both efficiency and effectiveness

Conclusion & Future Work

- CFM for feature-based recommendation
  - Outer product-based interaction cube
  - 3D CNN to explicitly learn high-order interactions
  - Attention-based feature pooling layer to reduce computational cost
- Future work
  - Improve efficiency
  - Residual learning

Reference

- [Rendle et al.,2010] Factorization machines. In ICDM.
- [Rendle et al.,2009] Bpr: Bayesian personalized ranking from implicit feedback. In UAI.
- [He, et al.,2017] Neural factorization machines for sparse predictive analytics. In SIGIR.
- [Guo et al.,2017] Deepfm: A factorization-machine based neural network for ctr prediction. In IJCAI.
- [He et al.,2018] Outer product-based neural collaborative filtering. In IJCAI.
- [Zhang et al.,2014] Explicit factor models for explainable recommendation based on phrase-level sentiment analysis. In SIGIR.

Thank you

Q&A