据的预测,所以 DeepFM 的深度部分采用 DNN 模型.

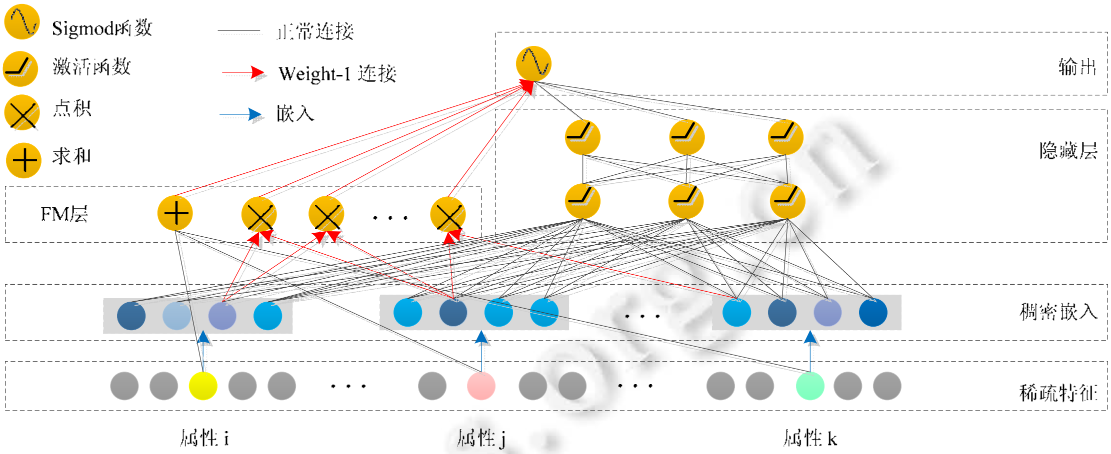

Fig.6 Architecture of DeepFM model

图 6 DeepFM 模型结构

3.5 NFM模型

NFM 模型[74]也采用了类似 Wide&Deep 的宽度和深度学习框架,其输出表示为

$$\hat{y}(x)=w_0+\sum_{i=1}^{n}w_ix_i+f(x) \tag{25}$$

其中,第 1、2 部分是线性回归模型,与 FM 一致;第 3 部分的 $f(x)$是模型的核心,主要对特征交互进行建模,是一个多层前向神经网络.图 7 所示为本文对该模型的结构描述(深度学习部分的输入是交互特征).

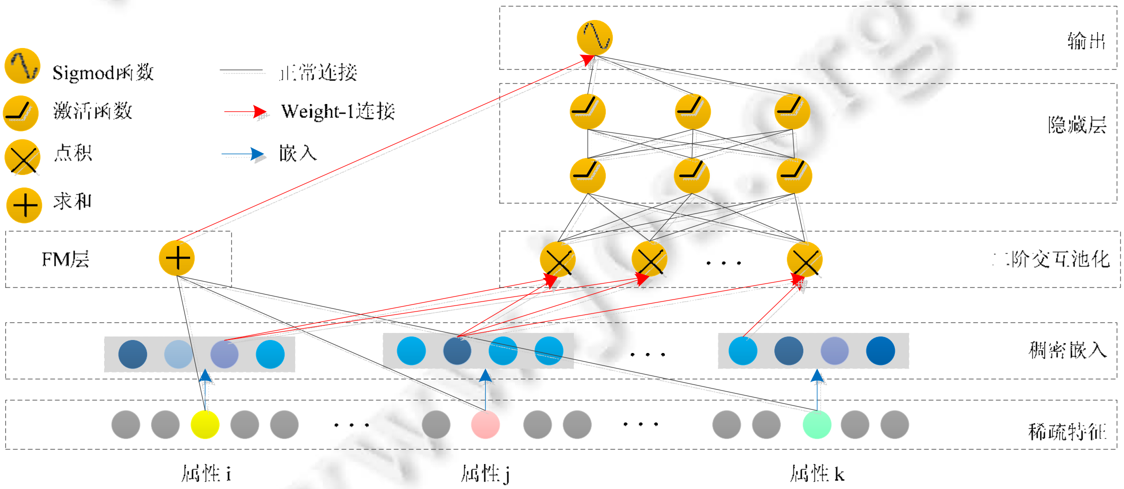

Fig.7 Architecture of NFM model

图 7 NFM 模型结构

3.6 宽度和深度学习模型集成方式分析

FNN 模型实现了 Wide&Deep 框架的深度部分,它采用 FM 对参数进行初始化,然后再深度学习.DeepFM 与 PNN 结构很相似,不同在于 FM 模型处理的属性被单独作为宽度部分.DeepFM 与 Wide&Deep 的不同在于:它把宽度部分的属性替换为 FM 处理后的属性,而且宽度部分与深度部分输入相同.NFM 模型也是基于 Wide&Deep

框架,主要通过输入特征的预处理来提高推荐效果.Deep&Cross 模型则是对文本信息进行处理,并且将 ResNet[75]应用到非图像领域.表 2 对这些模型的特征进行了综合比较.

Table 2 Comparison of depth learning model based on FM model and linear model

表 2 基于 FM 模型的深度学习模型比较

模型	1 阶特征	2 阶特征	高阶特征学习	高阶输入	共享输入
FNN	×	×	deep 网络	嵌入向量	√
Wide&Deep	√	少量,人工	deep 网络	嵌入向量	×
Deep&Cross	cross 网络	cross 网络	cross+deep 网络	嵌入向量	√
DeepFM	√	√	deep 网络	嵌入向量	√
NFM	√	×	deep 网络	2 阶特征	√

这些模型基本包括了传统模型(尤其是 FM)与深度学习模型的不同融合方式,有些是松耦合(两者的最后结果进行合并),有些是紧耦合(一个模型的输入依赖于另一个模型).实际应用中,如果要提高预测或推荐的精度,需要结合应用领域来增加特征工程.

为了深入理解不同模型及其不同结构的优势,本项目组提出了 DGFFM 模型.模型中,宽度部分采用 FFM 模型,深度部分采用 DenseNet 模型,并且在特征中添加了时间动态因子.因为 FNN 和 Wide&Deep 采用两种典型的框架,所以本文实现了两种 DGFFM 模型:DGFFM(W&D)模型采用 Wide&Deep 结构,DGFFM(FNN)采用 FNN 的结构.实验采用的硬件平台为 Inter® Core™ i7-7700 CPU@3.60GHz,65.86GB 内存,976GB 硬盘,64 位 Ubuntu 16.04 操作系统的工作站;编程语言为 Python,框架使用 TensorFlow;数据集为 MovieLens 1M 数据集(简写为 ml-1m,http://grouplens.org/datasets/movielens/1m/)和 Criteo 数据集,ml-1m(softmax 分类)中采用 RMSE 评价指标,Criteo(二分类问题)中采用 AUC 和 LogLoss 评价指标.模型比较结果见表 3,其中,FNN,DeepFM 和 DGFFM 的隐向量维度均设置为 20.DenseNet 部分输出通道设置:ml-1m 为[100,48,32],分别指模块 1 输出通道数、翻译层输出通道数和模块 2 输出通道数,Criteo 为[256,128,64].由于各种深度模型适用于不同的应用场合,所以在实验中没有测试每种模型的效果,只把所提出的 DGFFM 模型与 FNN 和 DeepFM(采用 Wide&Deep 框架)模型进行了对比.

Table 3 Accurace comparison of different models

表 3 模型精度比较

数据集	FNN	DeepFM	DGFFM(FNN)	DGFFM(W&D)
ml-1m	RMSE:0.7608	RMSE:0.7493	RMSE:0.7231	RMSE:0.7028
Criteo	AUC:0.7935 LogLoss:0.4589	AUC:0.7989 LogLoss:0.4503	AUC:0.8167 LogLoss:0.4264	AUC: 0.8185 LogLoss:0.4118

根据表 3 的实验结果,得出以下结论.

1) DeepFM 在两个数据集上均比 FNN 取得了更好的效果.在 ml-1m 上,DeepFM 的 RMSE 减少 1.5%;在 Criteo 上,LogLoss 减少 1.9%,AUC 提高 0.6%.这在一定程度上说明 Wide&Deep 具有结构优势;
2) 3 个模型中,DGFFM 均取得了最好的结果.其原因是 DGFFM 在宽度学习部分基于 FFM 模型,增加了时间因子以及其他的特征工程,而且 DenseNet 相对于标准 DNN 也具有一定优势,宽度和深度两部分优势相结合,进一步提高了模型的预测精度;
3) DGFFM(W&D)结果略好于 DGFFM(FNN),在 ml-1m 上,RMSE 减少 2.8%;在 Criteo 上,LogLoss 减少 3.4%,AUC 提高 0.2%.由于 Wide&Deep 结构中为了保证最终层数不变,深度部分初始输入采用的是偏置+一次项,因此 Wide&Deep 结构中的 DGFFM 比 FNN 结构中的 DGFFM 多了一部分,这可能对最终结果也造成了一些影响.但总体而言,Wide&Deep 结构略好于 FNN 结构.

4 FM 模型学习与分布式并行实现

4.1 FM模型的学习与优化

Rendle 等人采用 3 种学习方法训练 FM 模型:SGD、交替最小二乘法(alternating least-squares,简称 ALS)和马尔可夫蒙特卡洛(Markov chain Monte Carlo,简称 MCMC),这些都可以在 libFM 中找到源码[21].Bayer 等人提出了库 fastFM[76],实现了 FM 的回归、分类和排序,简化了 FM 的使用.从关系型数据设计矩阵会非常庞大,使得学习和预测变得缓慢或者从标准的机器学习算法不可行的角度,Rendle 等人针对关系型数据实现了 FM[77].针对 FM 中包含一个非凸优化问题,导致局部最小化,Blondel 等人提出基于核范式的 FM 的凸形式,采用双块坐标下降算法(two-block coordinate descent algorithm)优化学习[78].Yuan 等人提出两种优化策略——RankingFM(ranking factorization machine)和 LambdaFM(lambda factorization machine)优化 FM 模型[79].

Pan 等人针对广告交易数据的稀疏性,即存在大量零元素,可能会严重影响 FM 模型的性能,提出一种新的稀疏因子分解机(SFM)模型[80],其中使用拉普拉斯分布而不是传统的高斯分布来对参数进行建模,因为拉普拉斯分布可以更好地拟合更高比例的稀疏数据零元素.Saha 等人提出了 NPFM[81],它假设数据服从泊松分布(Poisson distribution),这对于建模和数据训练计算都非常有利;NPFM 作为一个非参数模型,会从数据本身发现最适合的隐因子数量.由于 FM 中用户、物品和上下文变量之间的交互被建模成它们各自隐因子特征的线性组合(linear combination),但是将用户、物品和上下文变量之间的交互限制成线性组合并不现实,为了解决这一限制,Nguyen 等人提出了高斯过程的因子分解机(Gaussian process factorization machine,简称 GPFM)模型,即:使用高斯过程的非线性概率算法来应对上下文感知推荐,可以被应用到隐式反馈数据和显式反馈数据集[82].一般的高斯处理回归的推断和学习算法都是关于数据集样本大小的立方级复杂度(cubic complexity),Huang 等人提出了 GGPFM(grid-based Gaussian processes factorization machine)模型捕捉用户与物品之间的非线性交互[83],将潜在特征(latent features)赋予网格结构(grid structures)降低模型复杂度.通常的学习和训练方法即可满足一般的应用需求,但是不同的应用背景对应不同的特征,模型学习和优化方式的调整可以提高精度,不过都需要通过实验来反复验证.

4.2 FM模型的并行实现

精度和效率是评价预测/推荐模型的两个重要指标.通过从宽度上改进模型以及从深度上与深度学习模型的集成,可以极大地提高模型的精度.FM 提供了线性的计算复杂度和有用的数据嵌入,但是当数据和特征规模增大时,模型的扩展代价非常高.FM 与深度学习集成后,大数据和模型扩展性问题更加严重.机器学习算法的独特性在于:(1) 迭代性,模型的更新需要循环迭代多次;(2) 容错性,每个循环中可能产生的错误不影响模型最终的收敛;(3) 参数收敛的非均匀性,模型中有些参数经过几次循环后不再改变,其他参数可能仍需要很长时间收敛.面对海量的数据加上复杂的数学运算,这些决定了分布式机器学习系统的特殊性.大数据的机器学习存在着许多挑战和机遇[84,85],通常会采用两种方案.

- 首先是数据并行方案.采用经典的主-从服务模式对训练数据进行划分,分布式存储到各个节点上,每个节点都运行着一个或多个模型训练进程,各自完成前向和后向的计算得到梯度;训练结束后,各节点把参数传递给主服务器进行参数的合并与更新,主服务器把更新后的参数再分发到各个节点,再次进行训练.通过多个节点并行训练来提高学习效率;
- 其次是结构并行方案.当模型巨大、单机内存不足时,将计算工作进行划分,即同一个大模型的不同部分交给不同节点负责(如多层网络的各个节点),不过,这样会产生很大的通信开销.结构并行相对数据并行更加复杂,不过开源框架如 TensorFlow 平台直接支持结构并行.

目前,主流的解决方案是使用分布式框架和并行计算模式,硬件方面则使用 GPU 和 TPU 等进行加速.以下将总结 FM 及其变体在提高效率方面的相关研究.

MapReduce 并行计算模式在大数据处理领域应用非常广泛,也可被用于提高 FM 的学习效率.Sun 等人实现了基于 MapReduce 的 SGD 算法用于 FM 模型的学习,主要通过数据并行来提高模型的学习效率.不过,

MapReduce 模式的特征也决定了其更适合处理数据并行[86].Yan 等人基于 spark 平台实现 FM 模型的学习,其核心思想与 MapReduce 类似[47].Knoll 等人采用参数服务器(parameter server,简称 PS)为 FM 提出一种分布式的 SG 算法[22].PS 是一个算法的计算引擎,其计算由两组分开的计算机完成:服务器(server)和工作者(worker). server 用于管理和更新模型的参数,worker 处理训练数据,任务调度器和资源管理器负责控制数据流.Li 等人也采用 PS,通过一个依赖图(dependency graph,简称 DAG)提供了灵活的数据一致性模型[87].Zhong 等人在参数服务器上实现了分布式的 FM,即 DiFacto,采用自适应的内存限制和频度自适应的正则化机制,基于数据和模型统计来执行细粒度的控制,并在多台机器分发 DiFacto[88].Li 等人提出一个新的系统框架,集成了参数服务器和 MapReduce 模式.通过 MapReduce 实现数据并行,通过 PS 实现模型并行,并解决了通信开销问题和参数更新冲突问题[89].机器学习离不开分布式并行计算框架和 GPU 等硬件的支持.

5 FM 模型研究展望

“互联网+”的发展以及大数据技术正在开启一个全新认知的大数据时代,FM 模型是目前预测/推荐领域研究和应用最广泛的模型之一.图 8 所示大数据环境下预测/推荐系统的框架及其所面临的问题.

- 问题 1:多源异构数据带来特征表示的多样性和复杂性.尤其是视觉数据,其特征维度高且数量大,传统推荐主要关注非视觉文本数据及其交互,对非视觉和视觉特征的融合是新型推荐系统建立的基础;
- 问题 2:现有推荐模型对于特征进化趋势缺乏表示.用户偏好与物品特征都会随着时间而发生变化,对这些变化趋势进行合理建模将会提升推荐效果,动态特征的提取与表示是构建动态推荐模型的关键;
- 问题 3:传统模型与深度学习框架适用于不同领域,现有的融合方法不论是共享输入还是独立输入,两部分都是松耦合,两者的结合主要用于增加推荐的多样性.为了发挥共同优势,需要研究模型的集成策略;
- 问题 4:大数据要求模型和训练算法具有高扩展性和高效率,尽管通用分布式计算框架提供了并行处理支持,但两类不同的学习方式具有不同的要求,相关并行处理方法和关键技术还需进一步研究.

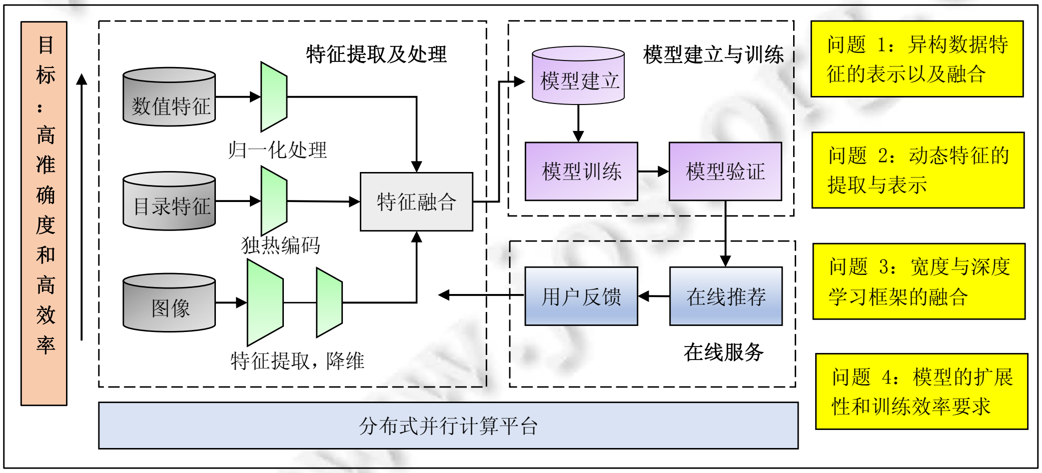

Fig.8 Framework of prediction & recommendation system and the problems faced by big data environment

图 8 预测/推荐系统框架及大数据环境下所面对的问题

通过对已有 FM 的相关研究和应用进行分析,我们认为,目前工作可以从以下两方面进行深入.

5.1 时间动态性

动态建模是推荐系统面临的挑战之一.Koren 将时间动态性应用到矩阵奇异值分解模型 SVD 中,通过提取非视觉特征的时间动态因子,分别对用户偏置和物品偏置进行动态建模,取得了较好的推荐效果[11].He 等人同

时对非视觉特征和视觉特征进行轻量级的时间建模,虽然没有对变化趋势进行细分,但也极大地改善了推荐结果[90].谷歌将流行趋势划分为 6 类:持续上升、季节上升、突然上升、持续下降、季节下降和突然下降,基本囊括时尚物品的所有变化特性.用户行为变化趋势和物品变化趋势有较大差异,通常短期内物品的流行趋势变化不明显,研究不同特征的不同变化趋势对于构建推荐模型具有重要意义[91].相关研究属于特征工程领域,即在原有的属性中添加时间因子.FM 模型中可以归纳为两类时间动态性:偏置动态性和特征动态性.

(1) 偏置动态性.

在推荐系统中,又分为用户的偏置动态性和物品的偏置动态性.设置用户偏置动态性的原因在于:用户对物品的评分习惯可能会随着时间而发生变化.例如,用户 Aphro 过去倾向于给电影《秒速五厘米》评 9 分,现在她对于动画片的狂热减退,只会评 8 分.同样地,设置物品的偏置动态性的原因在于:随着时间的推移,物品的人气会随之变化.以电影《战狼 2》为例,上映 4 小时的票房达 9 741 万,接着,凭借演员精湛的演绎、电影较好的口碑以及网络话题的引燃而迅速火热起来,上映 10 天,票房便突破 31 亿.偏置动态性的表示可以在用户偏置 b_u 中添加如下时间函数:

$$b(t)=b+\tau(t) \tag{26}$$

其中,b 是静态偏置,$\tau(t)$是一个时间函数.对$\tau(t)$进行建模,常用方式是建立简单的时间线性模型,如 timeSVD++.然而,实际应用场景的表现通常是非线性的,甚至可能是无规则的,很难用确定的公式表示.

(2) 特征动态性.

对于属性 $x_i(1\leqslant i\leqslant n)$,可以细分成随时间变化的动态属性和保持稳定的静态属性,静态属性不需额外处理,动态属性则可添加时间变化函数,调整为

$$x_i(t)=x_i f(t) \tag{27}$$

其中,$f(\cdot)$是非线性激活函数.这些偏置函数和属性函数可以加入到 FM 模型中,对其静态偏置和属性进行调整.

5.2 视觉和非视觉属性融合

传统推荐方法与模型主要关注非视觉文本特征及其交互,如用户与物品的固有特征描述、物品的星级评分、用户的购买历史、书签、浏览日志、查询模式、鼠标活动等.随着机器视觉领域中深度学习的广泛应用,图像特征开始被关注,高维视觉特征也作为预测/推荐模型属性的一部分.视觉特征容易获得且描述准确,因此,如何提取高维视觉特征以及如何把非视觉低维特征和视觉高维特征进行有效融合,成为目前推荐领域的研究热点[92−94],并在 RecSys 2017 会议中作为重要主题列出.

deep CNN 模型最近被成功应用于对象检测、图像匹配等领域,相关研究已证明:基于海量数据训练的 deep CNN 模型可以精确应用于其他数据集,在新的数据集中仍然能够产生很好的效果.假设 f_i 表示物品 i 的原始视觉特征向量,维度为 D 维(可以通过 AlexNet,ResNet 等预先训练好,比如取 AlexNet 的第 2 个全连接层 FC7 的输出作为原始视觉特征向量),那么视觉特征θ_i可以按如下方式建模:

$$\theta_i=\boldsymbol{E}^T f_i \tag{28}$$

其中,E 是一个维度为 $D\times F$ 矩阵.此时,θ_i的特征维度为 $F(F<<D)$,从而达到降维的目的.两个物品之间的视觉关系可以表示为

$$\theta_{i,j}=\boldsymbol{E}^T(f_i-f_j) \tag{29}$$

这种低秩嵌入方法仅仅能捕获两个物品是否关联,关联的原因则不能表达.实际应用中,物品之间的关联关系可能体现为多种原因,如一件 T-恤和一条短裙搭配合适的原因可能是颜色、质地或者款式等.为了解决这个问题,可以考虑采用多重嵌入,两个物品之间的关系可以表示为

$$\theta_{i,j}=\boldsymbol{E}_0^T f_i-\boldsymbol{E}_k^T f_j \tag{30}$$

其中,E_0 把物品 i 对应到一个参考点,$\boldsymbol{E}_0^T f_i$ 对应到一个嵌入空间,$E_k(k=1,2,\ldots,N)$表示与物品 j 的潜在匹配.传统的 FM 主要面向文本等非视觉特征,如果把视觉特征也融入到 FM 模型中,这样 FM 模型的应用将更加广泛.

6 结束语

在预测/推荐系统领域,FM 模型被广泛研究与应用.没有万能的模型,不同业务场景对模型的输入特征、处理逻辑和输出类别会有不同要求.本文从宽度扩展和深度扩展视角对 FM 模型及其变体的研究进行综述,希望能够提供不同的思路,为应用提供不同的选择方案.从对国内外高水平期刊及会议上的文献分析可以看出:将传统 FM 模型与深度学习模型相结合、将视觉与非视觉特征进行融合是研究热点.现有相关工作可以从以下几点进行深入:(1) 研究非视觉特征和视觉特征的融合,目前的相关研究缺乏对用户行为和物品变化趋势的差异化和细粒度处理;(2) 宽度和深度两种学习方式的融合目前主要用于增加推荐的多样性,其耦合方式以及对精度和效率的影响还需进一步研究.

References:

[1] Bobadilla J, Ortega F, Hernando A, Gutierrez A. Recommender systems survey. Knowledge-Based Systems, 2013,46(1):109−132.

[2] Uribe G, Carlos A, Hunt N. The netflix recommender system: algorithms, business value, and innovation. ACM Trans. on Management Information Systems, 2016,6(4):1−19.

[3] Covington P, Adams J, Sargin E. Deep neural networks for youtube recommendations. In: Proc. of the Conf. on Recommender Systems. New York: ACM Press, 2016. 191−198.

[4] Okura S, Tagami Y, Ono S, Tajima A. Embedding-based news recommendation for millions of users. In: Proc. of the 23rd ACM SIGKDD Conf. on Knowledge Discovery and Data Mining. New York: ACM Press, 2017. 1933−1942.

[5] Kazai G, Yusof I, Clarke D. Personalised news and blog recommendations based on user location, facebook and Twitter user profiling. In: Proc. of the 39th ACM SIGIR Conf. on Research and Development in Information Retrieval. New York: ACM Press, 2016. 1129−1132.

[6] Huang L, Lin CJ, He J, Liu HY, Du XY. Diversified mobile app recommendation combining topic model and collaborative filtering. Ruan Jian Xue Bao/Journal of Software, 2017,28(3):708−720 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/5163.htm [doi: 10.13328/j.cnki.jos.005163]

[7] Cheng HT, Koc L, Harmsen J, Shaked T, Chandra T, Aradhye H, Anderson G, Corrado G, Chai W, Ispir M, Anil R, Haque Z, Hong LC, Jain V, Liu XB, Shah H. Wide & deep learning for recommender systems. In: Proc. of the 1st Workshop on Deep Learning for Recommender Systems. New York: ACM Press, 2016. 7−10.

[8] Yan CR, Zhang QL, Zhao X, Huang YF. Method of bayesian probabilistic matrix factorization based on generalized Gaussian distribution. Journal of Computer Research and Development, 2016,53(12):2793−2800 (in Chinese with English abstract).

[9] Zhao QB, Zhou GX, Zhang LQ, Cichocki A, Amari S. Bayesian robust tensor factorization for incomplete multiway data. IEEE Trans. on Neural Networks and Learning Systems, 2016,27(4):736−748.

[10] Koren Y. Factorization meets the neighborhood: A multifaceted collaborative filtering model. In: Proc. of the 14th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining. New York: ACM Press, 2008. 426−434.

[11] Koren Y. Collaborative filtering with temporal dynamics. In: Proc. of the 15th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM Press, 2009. 447−456.

[12] Rendle S. Factorization machines. In: Proc. of the 10th IEEE Int'l Conf. on Data Mining. Piscataway: IEEE, 2010. 995−1000.

[13] Juan YC, Zhuang Y, Chin WS, Lin CJ. Field-aware factorization machines for CTR prediction. In: Proc. of the 10th ACM Conf. on Recommender Systems. New York: ACM Press, 2016. 43−50.

[14] Ta A. Factorization machines with follow-the-regularized-leader for CTR prediction in display advertising. In: Proc. of the 2015 IEEE Int'l Conf. on Big Data. Piscataway: IEEE, 2015. 2889−2891.

[15] Resnick P, Iacovou N, Suchak M, Bergstrom P, Riedl J. GroupLens: An open architecture for collaborative filtering of netnews. In: Proc. of the ACM Conf. on Computer Supported Cooperative Work. New York: ACM Press, 1994. 175−186.

[16] Herlocker JL, Konstan JA, Borchers A. An algorithmic framework for performing collaborative filtering. In: Proc. of the 22nd Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. New York: ACM Press, 1999. 230−237.

[17] Meng XW, Liu SD, Zhang YJ, Hu X. Research on social recommender systems. Ruan Jian Xue Bao/Journal of Software, 2015, 26(6):1356−1372 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4831.htm [doi: 10.13328/j.cnki.jos.004831]

[18] Meng XW, Chen C, Zhang YJ. A survey of mobile news recommend techniques and applications. Chinese Journal of Computers, 2016,39(4):685−703 (in Chinese with English abstract).

[19] Hong H, Pradhan B, Sameen MI, Chen W, Xu C. Spatial prediction of rotational landslide using geographically weighted regression, logistic regression, and support vector machine models in Xing Guo area (China). Geomatics, Natural Hazards and Risk, 2017,8(2):1997−2022.

[20] Chang YW, Hsieh CJ, Chang KW, Ringgaard M, Lin CJ. Training and testing low-degree polynomial data mappings via linear SVM. Journal of Machine Learning Research, 2010,11(4):1471−1490.

[21] Rendle S. Factorization machines with libFM. ACM Trans. on Intelligent Systems and Technology, 2012,3(3):57.

[22] Knoll J. Recommending with higher-order factorization machines. In: Proc. of the Int'l Conf. on Innovative Techniques and Applications of Artificial Intelligence. Berlin: Springer-Verlag, 2016. 103−116.

[23] Blondel M, Fujino A, Ueda N, Ueda N, Ishihata M. Higher-order factorization machines. In: Proc. of the 30th Conf. on Neural Information Processing Systems. Berkeley: USENIX, 2016. 3351−3359.

[24] Prillo S. An elementary view on factorization machines. In: Proc. of the 11st ACM Conf. on Recommender Systems. New York: ACM Press, 2017. 179−183.

[25] Knoll J, Köckritz D, Groß R. Markov random walk vs. higher-order factorization machines: A comparison of state-of-the-art recommender algorithms. In: Proc. of the Int'l Conf. on Innovations for Community Services. Berlin: Springer-Verlag, 2017. 87−103.

[26] Yurochkin M, Nguyen XL. Multi-way interacting regression via factorization machines. In: Proc. of the 31st Conf. on Neural Information Processing Systems. Berkeley: USENIX, 2017. 2595−2603.

[27] Pan J, Xu J, Ruiz AL, Zhao W, Pan S, Sun Y, Lu Q. Field-weighted factorization machines for click-through rate prediction in display advertising. In: Proc. of the 2018 World Wide Web Conf. on World Wide Web. Berlin: Springer-Verlag, 2018. 1349−1357.

[28] Zhao Y, Mansouri K, Yang Y, Mi ZQ. Rating prediction using category weight factorization machine in bigdata environment. In: Proc. of the IEEE Int'l Conf. on Communication Workshop. Piscataway: IEEE, 2015. 1909−1913.

[29] Wang S, Li C, Zhao K, Chen H. Learning to context-aware recommend with hierarchical factorization machines. Information Sciences, 2017,409:121−138.

[30] Guo R, Alvari H, Shakaria P. Strongly hierarchical factorization machines and anova kernel regression. In: Proc. of the 2018 SIAM Int'l Conf. on Data Mining. Philadelphia: SIAM, 2018. 729−737.

[31] Oentaryo RJ, Lim EP, Low JW, Lo D. Predicting response in mobile advertising with hierarchical importance-aware factorization machine. In: Proc. of the 7th ACM Int'l Conf. on Web Search and Data Mining. New York: ACM Press, 2014. 123−132.

[32] Yuan FJ, Guo GB, Jose JM, Chen L, Yu H, Zhang W. BoostFM: Boosted factorization machines for top-*n* feature-based recommendation. In: Proc. of the 22nd Int'l Conf. on Intelligent User Interfaces. New York: ACM Press, 2017. 45−54.

[33] Yan P, Zhou X, Duan Y. E-Commerce item recommendation based on field-aware factorization machine. In: Proc. of the 2015 Int'l ACM Recommender Systems Challenge. New York: ACM Press, 2015..

[34] Hong L, Doumith AS, Davison BD. Co-Factorization machines: Modeling user interests and predicting individual decisions in Twitter. In: Proc. of the 6th ACM Int'l Conf. on Web Search and Data Mining. New York: ACM Press, 2013. 557−566.

[35] Leksin V, Ostapets A. Job recommendation based on factorization machine and topic modelling. In: Proc. of the Recommender Systems Challenge. New York: ACM Press, 2016..

[36] Blondel M, Niculae V, Otsuka T, Ueda N. Multi-Output polynomial networks and factorization machines. In: Proc. of the 31st Conf. on Neural Information Processing Systems. Berkeley: USENIX, 2017. 3351−3361.

[37] Wang S, Du C, Zhao K, Li C, Li Y, Zheng Y, Wang Z, Chen H. Random partition factorization machines for context-aware recommendations. In: Proc. of the Int'l Conf. on Web-Age Information Management. Berlin: Springer-Verlag, 2016. 219−230.

[38] Pijnenburg M, Kowalczyk W. Extending logistic regression models with factorization machines. In: Proc. of the Int'l Symp. on Methodologies for Intelligent Systems. Berlin: Springer-Verlag, 2017. 323−332.

[39] Loni B, Said A, Larson M, Hanjalic A. 'Free lunch' enhancement for collaborative filtering with factorization machines. In: Proc. of the 8th ACM Conf. on Recommender Systems. New York: ACM Press, 2014. 281−284.

[40] Cheng C, Xia F, Zhang T, King L, Lyu M. Gradient boosting factorization machines. In: Proc. of the 8th ACM Conf. on Recommender Systems. New York: ACM Press, 2014. 265−272.

[41] Xu J, Lin K, Tan PN, Zhou J. Synergies that matter: Efficient interaction selection via sparse factorization machine. In: Proc. of the 2016 SIAM Int'l Conf. on Data Mining. Philadelphia: SIAM, 2016. 108−116.

[42] Selsaas LR, Agrawal B, Rong C, Wiktorski T. AFFM: Auto feature engineering in field-aware factorization machines for predictive analytics. In: Proc. of the IEEE Int'l Conf. on Data Mining Workshop. Piscataway: IEEE, 2015. 1705−1709.

[43] Punjabi S, Bhatt P. Robust factorization machines for user response prediction. In: Proc. of the 2018 World Wide Web Conf. on World Wide Web. Berlin: Springer-Verlag, 2018. 669−678.

[44] Lu CT, He L, Shao W, Cao B, Yu PS. Multilinear factorization machines for multi-task multi-view learning. In: Proc. of the 10th ACM Int'l Conf. on Web Search and Data Mining. New York: ACM Press, 2017. 701−709.

[45] Liu C, Zhang T, Zhao P, Zhou J, Sun JL. Locally linear factorization machines. In: Proc. of the 26th Int'l Joint Conf. on Artificial Intelligence. Menlo Park: AAAI, 2017. 2294−2300.

[46] Yan CR, Zhang QL, Zhao X, Huang YF. An intelligent field-aware factorization machine mode. In: Proc. of the Int'l Conf. on Database Systems for Advanced Applications. Berlin: Springer-Verlag, 2017. 309−323.

[47] Ding Y, Wang D, Xin X, Li GQ, Sun D, Zeng XZ. SCFM: Social and crowdsourcing factorization machines for recommendation. Applied Soft Computing, 2018,66:548−556.

[48] Zhou J, Wang D, Ding Y, Yin L. SocialFM: A social recommender system with factorization machines. In: Proc. of the Int'l Conf. on Web-Age Information Management. Berlin: Springer-Verlag, 2016. 286−297.

[49] Rendle S. Social network and click-through prediction with factorization machines. In: Proc. of the KDD-Cup Workshop. New York: ACM Press, 2012. 113.

[50] Chen CM, Chen HP, Tsai MF, Yang YH. Leverage item popularity and recommendation quality via cost-sensitive factorization machines. In: Proc. of the 2014 IEEE Int'l Conf. on Data Mining Workshop. Piscataway: IEEE 2014. 1158−1162.

[51] Qiang R, Liang F, Yang J. Exploiting ranking factorization machines for microblog retrieval. In: Proc. of the 22nd ACM Int'l Conf. on Information & Knowledge Management. New York: ACM Press, 2013. 1783−1788.

[52] Xu Y, Tang Q, Hou LZ, Li M. Decision model for market of performing arts with factorization machine. Journal of Shanghai Jiaotong University (Science), 2018,23(1):74−84.

[53] Juan Y, Lefortier D, Chapelle O. Field-Aware factorization machines in a real-world online advertising system. In: Proc. of the 26th Int'l Conf. on World Wide Web Companion. Berlin: Springer-Verlag, 2017. 680−688.

[54] Chen C, Hou C, Xiao J, Yuan XJ. Purchase behavior prediction in e-commerce with factorization machines. IEICE Trans. on Information and Systems, 2016,99(1):270−274.

[55] Wang Y, Shang W, Li Z. The application of factorization machines in user behavior prediction. In: Proc. of the 15th Int'l Conf. on Computer and Information Science. Piscataway: IEEE, 2016. 1−4.

[56] Cao B, Shi M, Liu XF, Liu JX, Tang Md. Using relational topic model and factorization machines to recommend Web apis for mashup creation. In: Proc. of the Asia-Pacific Services Computing Conf. Berlin: Springer-Verlag, 2016. 391−407.

[57] Wu Y, Xie F, Chen L, Chen C, Zheng Z. An embedding based factorization machine approach for Web service qos prediction. In: Proc. of the Int'l Conf. on Service-Oriented Computing. Berlin: Springer-Verlag, 2017. 272−286.

[58] Tang MD, Zhang TT, Yang YT, Zheng ZB, Cao BQ. QoS-Aware Web service recommendation based on factorization machines. Chinese Journal of Computers, 2018,41(6):1300−1313 (in Chinese with English abstract).

[59] Chen C, Wu D, Hou CY, Yuan XJ. Exploiting social media for stock market prediction with factorization machine. In: Proc. of the 2014 IEEE/WIC/ACM Int'l Joint Conf. on Web Intelligence and Intelligent Agent Technologies. Piscataway: IEEE, 2014. 142−149.

[60] Yamada M, Lian W, Goyal A, Chen JH, Wimalaweane H, Khan SA. Convex factorization machine for toxicogenomics prediction. In: Proc. of the 23rd ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM Press, 2017. 1215−1224.

[61] Zhu G, Li L. Factorization machine based business credit scoring by leveraging internet data. In: Proc. of the Asia-Pacific Web Technologies and Applications. Berlin: Springer-Verlag, 2016. 565−569.

[62] Sun LJ, Fan JF, Yang WQ, Shi YH. Application of factorization machine in mobile App recommendation based on deep packet inspections. Journal of Computer Applications, 2016,36(2):307−310 (in Chinese with English abstract).

[63] Zhu M, Aggarwal CC, Ma S. Outlier detection in sparse data with factorization machines. In: Proc. of the 2017 ACM Conf. on Information and Knowledge Management. New York: ACM Press, 2017. 817−826.

[64] Zheng L, Noroozi V, Yu PS. Joint deep modeling of users and items using reviews for recommendation. In: Proc. of the 10th ACM Int'l Conf. on Web Search and Data Mining. New York: ACM Press, 2017. 425−434.

[65] Chen J, Sun B, Li H, Lu HT, Hua XS. Deep CTR prediction in display advertising. In: Proc. of the 24th ACM Int'l Conf. on Multimedia. New York: ACM Press, 2016. 811−820.

[66] Bracher C, Heinz S, Vollgraf R. Fashion DNA: Merging content and sales data for recommendation and article mapping. In: Proc. of the 22nd ACM SIGKDD Conf. on Knowledge Discovery and Data Mining, Fashion Workshop. New York: ACM Press, 2016.

[67] Zhang S, Yao L, Sun A. Deep learning based recommender system: A survey and new perspectives. ACM Journal on Computing and Cultural Heritage, 2017,1(1):35.

[68] Razavian AS, Azizpour H, Sullivan J, Carlsson S. CNN features off-the-shelf: An astounding baseline for recognition. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE, 2014. 806−813.

[69] Wang R, Fu B, Fu G, Wang ML. Deep & cross network for ad click predictions. In: Proc. of the 17th Knowledge Discovery and Data Mining. New York: ACM Press, 2017.

[70] Zhang W, Du T, Wang J. Deep learning over multi-field categorical data. In: Proc. of the European Conf. on Information Retrieval. Berlin: Springer-Verlag, 2016. 45−57.

[71] Qu Y, Cai H, Ren K, Zhang WN, Yu Y. Product-Based neural networks for user response prediction. In: Proc. of the 2016 IEEE 16th Int'l Conf. on Data Mining. Piscataway: IEEE, 2016. 1149−1154.

[72] Shan Y, Hoens TR, Jiao J, Wang HJ, Yu D, Mao JC. Deep crossing: Web-scale modeling without manually crafted combinatorial features. In: Proc. of the 22nd Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM Press, 2016. 255−262.

[73] Guo H, Tang R, Ye Y, Li ZG, He XQ. DeepFM: A factorization-machine based neural network for CTR prediction. In: Proc. of the 26th Int'l Joint Conf. on Artificial Intelligence. Berlin: Springer-Verlag, 2017. 1725−1731.

[74] He X, Chua TS. Neural factorization machines for sparse predictive analytics. In: Proc. of the 40th Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. New York: ACM Press, 2017. 355−364.

[75] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016. 770−778.

[76] Bayer I. FastFM: A library for factorization machines. Journal of Machine Learning Research, 2016,17:1−5.

[77] Rendle S. Scaling factorization machines to relational data. In: Proc. of the Very Large Data Bases Conf. Endowment. Trondheim: VLDB, 2013,6(5):337−348.

[78] Blondel M, Fujino A, Ueda N. Convex factorization machines. In: Proc. of the Joint European Conf. on Machine Learning and Knowledge Discovery in Databases. Berlin: Springer-Verlag, 2015. 19−35.

[79] Yuan F, Guo G, Jose JM, Chen L, Yu HT. Optimizing factorization machines for top-*n* context-aware recommendations. In: Proc. of the Web Information Systems Engineering (WISE 2016). Berlin: Springer-Verlag, 2016. 278−293.

[80] Pan Z, Chen E, Liu Q, Xu T, Ma HP, Lin HJ. Sparse factorization machines for click-through rate prediction. In: Proc. of the 2016 IEEE 16th Int'l Conf. on Data Mining. Piscataway: IEEE, 2016. 400−409.

[81] Saha A, Acharya A, Ravindran B, Ghosh J. Nonparametric poisson factorization machine. In: Proc. of the 2015 IEEE Int'l Conf. on Data Mining. Piscataway: IEEE, 2015. 967−972.

[82] Nguyen TV, Karatzoglou A, Baltrunas L. Gaussian process factorization machines for context-aware recommendations. In: Proc. of the 37th Int'l ACM SIGIR Conf. on Research & Development in Information Retrieval. New York: ACM Press, 2014. 63−72.

[83] Huang X, Yang Y, Bao X. Grid-Based Gaussian processes factorization machine for recommender systems. In: Proc. of the 9th Int'l Conf. on Machine Learning and Computing. New York: ACM Press, 2017. 92−97.

[84] Rendle S, Fetterly D, Shekita EJ, Su B. Robust large-scale machine learning in the cloud. In: Proc. of the 22nd ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM Press, 2016. 1125−1134.

[85] Zhou L, Pan S, Wang J, Vasilakos AV. Machine learning on big data: Opportunities and challenges. Neurocomputing. 2017,237: 350–361.

[86] Sun H, Wang W, Shi Z. Parallel factorization machine recommended algorithm based on mapreduce. In: Proc. of the 2014 10th Int'l Conf. on Semantics, Knowledge and Grids. Piscataway: IEEE, 2014. 120–123.

[87] Li M, Liu Z, Smola AJ, Wang YX. Difacto: Distributed factorization machines. In: Proc. of the 9th ACM Int'l Conf. on Web Search and Data Mining. New York: ACM Press, 2016. 377–386.

[88] Zhong E, Shi Y, Liu N, Rajan SJ. Scaling factorization machines with parameter server. In: Proc. of the 25th ACM Int'l Conf. on Information and Knowledge Management. New York: ACM Press, 2016. 1583–1592.

[89] Li M, Andersen DG, Park JW, Smola AJ, Ahmed A. Scaling distributed machine learning with the parameter server. In: Proc. of the 11th USENIX Symp. on Operating Systems Design and Implementation. Berkeley: USENIX, 2014. 583–598.

[90] He R, Mcauley J. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In: Proc. of the 25th Int'l Conf. on World Wide Web. Berlin: Springer-Verlag, 2016. 507–517.

[91] Wang Y, Ouyang H, Deng H, Chang Y. Learning online trends for interactive query auto-completion. IEEE Trans. on Knowledge and Data Engineering, 2017,29(11):2442–2454.

[92] Saha A, Raykar VC, Khapra M. Joint multi-modal representations for e-commerce catalog search driven by visual attributes. In: Proc. of the 22nd ACM SIGKDD Conf. on Knowledge Discovery and Data Mining. New York: ACM Press, 2016. 13–17.

[93] He R, McAuley J. VBPR: Visual bayesian personalized ranking from implicit feedback. In: Proc. of the 30th AAAI Conf. on Artificial Intelligence. Menlo Park: AAAI, 2016. 144–150.

[94] He R, Fang C, Wang Z, AcAuley J. Vista: A visually, socially, and temporally-aware model for artistic recommendation. In: Proc. of the 10th ACM Conf. on Recommender Systems. New York: ACM Press, 2016. 309–316.

附中文参考文献:

[6] 黄璐,林川杰,何军,刘红岩,杜小勇.融合主题模型和协同过滤的多样化移动应用推荐.软件学报,2017,28(3):708–720. http://www.jos.org.cn/1000-9825/5163.htm [doi: 10.13328/j.cnki.jos.005163]

[8] 燕彩蓉,张青龙,赵雪,黄永锋.基于广义高斯分布的贝叶斯概率矩阵分解方法.计算机研究与发展,2016,53(12):2793–2800.

[17] 孟祥武,刘树栋,张玉洁,胡勋.社会化推荐系统研究.软件学报,2015,26(6):1356–1372. http://www.jos.org.cn/1000-9825/4831.htm [doi: 10.13328/j.cnki.jos.004831]

[18] 孟祥武,陈诚,张玉洁.移动新闻推荐技术及其应用研究综述.计算机学报,2016,39(4):685–703.

[58] 唐明董,张婷婷,杨亚涛,郑子彬,曹步清.基于因子分解机的质量感知 Web 服务推荐方法.计算机学报,2018,41(6):1300–1313.

[62] 孙良君,范剑锋,杨婉琪,史颖欢.因子分解机算法在基于深度数据包检测的手机应用推荐中的应用.计算机应用,2016,36(2): 307–310.

燕彩蓉(1978－),女,湖北仙桃人,博士,副教授,CCF 专业会员,主要研究领域为云计算,大数据,机器学习.

张青龙(1990－),男,博士生,主要研究领域为推荐系统,机器学习.

周灵杰(1994－),男,学士,主要研究领域为图像处理,推荐算法,深度学习.

李晓林(1973－),男,博士,教授,博士生导师,主要研究领域为深度学习,云计算,大数据.