

基于高阶偏差的因子分解机推荐算法*

王子豪, 徐桂琼

(上海大学 管理学院, 上海 200444)

摘要: 在推荐系统中, 因评分尺度差异而造成的偏差问题一直影响着协同过滤算法的预测准确性。其中针对矩阵因子分解算法中的偏差问题, 提出一种基于高阶偏差的因子分解机算法。该算法首先按照评分偏差的现实特征对用户和项目进行划分, 再将偏差类别作为辅助特征集成到因子分解机中, 实现了评分预测中不同偏差用户、项目的高阶交互。在 MovieLens 数据集上的实验结果表明, 相比传统矩阵因子分解算法, 提出的算法具有更低的预测误差, 体现了其更好的推荐性能。

关键词: 推荐系统; 矩阵因子分解; 因子分解机; 评分偏差

中图分类号: TP181

文献标志码: A

文章编号: 1001-3695(2017)02-339-04

doi:10.3969/j.issn.1001-3695.2017.02.004

High-order biased factorization machine recommender algorithm

Wang Zihao, Xu Guiqiong

(School of Management, Shanghai University, Shanghai 200444, China)

Abstract: In recommender system, bias problem caused by different rating scales has always effected the predict precision of collaborative filtering. Concerning this bias problem of matrix factorization, this paper proposed a high-order biased factorization machine recommender algorithm. Firstly, it grouped users and items by their rating bias feature from real world, then integrated them into the factorization machine, which provided the high-order interactions between the different biased users and items. The experimental results on MovieLens datasets demonstrate that the proposed algorithm has lower prediction error than other traditional matrix factorization algorithms, which shows its better recommender performs.

Key words: recommender system; matrix factorization; factorization machine; rating bias

随着研究和实践的不断深入, 推荐系统已成为互联网中缓解信息过载问题的有效手段, 并已广泛地应用在信息服务领域。推荐系统利用评分等历史行为刻画用户的兴趣偏好, 并主动将符合其兴趣偏好的项目推荐给用户, 从而避免了用户在其他海量无关内容中的时间与精力消耗。

然而, 当前推荐系统仍然面临许多问题和挑战^[1]。除了常见的稀疏性、冷启动和实时性等问题之外, 因用户评分尺度不同和项目质量差异等因素造成的评分偏差问题也是影响推荐系统预测准确性的因素之一。传统的近邻协同过滤算法为了消除评分偏差问题, 采用皮尔逊相似度代替余弦相似度来计算用户和项目的相似性^[2-4]。但文献[5]指出这种评分偏差问题仍会导致近邻协同过滤算法将低相似度的用户选为邻居, 从而影响推荐的准确度和覆盖率。矩阵因子分解(matrix factorization, MF)^[6]是一种基于模型的协同过滤算法, 它相对于近邻协同过滤算法具有较高的准确度和应对稀疏性问题的能力, 因此吸引了越来越多的关注^[7-10]。

针对矩阵因子分解算法中评分偏差问题, 近年来国内外学者尝试了不同的方法, 大致可以分为两种: a) 构建偏差模式从而消除偏差, 如文献[11]将评分偏差作为一种高斯噪声从其概率模型中消除, 文献[12]根据用户评分特征提取出六种评分模式, 再通过缩放量表的方法消除不同评分模式中的偏差; b) 添加偏差项来平衡偏差, 如文献[13]在基本矩阵因子分解的基础上引入用户与项目的偏差项, 文献[14]进一步从项目

类别、用户会话和项目时间动态三个角度拓展了引入的偏差项。其中第二种方法具有更好的适用性和拓展性, 但目前的偏差项仅是一些简单的线性组合, 无法描述复杂现实背景下的评分偏差问题。

因子分解机(factorization machine, FM)^[15,16]是一种拓展的矩阵因子分解算法。它通过对输入特征的两两因子分解, 提供了相比线性模型更高阶的交互关系建模, 极大地提升了预测准确性。目前因子分解机已经被广泛应用在上下文感知推荐和社交推荐等领域中^[17-19]。本文通过分析用户与项目评分偏差问题的现实特征, 提出一种基于偏差的用户与项目划分方法, 并利用因子分解机构建评分偏差的高阶交互关系, 改进矩阵因子分解算法的评分预测。在 MovieLens 数据集上的实验表明, 本文提出的基于高阶偏差的因子分解机推荐算法(HOBFM)具有更好的预测准确性和推荐性能。

1 相关工作

1.1 基于偏差的矩阵因子分解算法

假设用户评分矩阵为 $R(m, n)$, 其中 m, n 分别表示用户数与项目数, 矩阵中的元素 R_{ij} 表示用户 i 对项目 j 的评分。基本矩阵因子分解算法(basic MF)^[6]的核心思想是从矩阵 R 中提取出一组潜在因子向量 u 和 v , 然后用对应的用户潜在因子向量 u_i 与项目潜在因子向量 v_j 的内积表示用户 i 对该项目 j 的

收稿日期: 2016-01-28; 修回日期: 2016-03-22 基金项目: 国家自然科学基金资助项目(11201290, 61104042)

作者简介: 王子豪(1992-), 男, 浙江温州人, 硕士研究生, 主要研究方向为个性化推荐、数据挖掘(wzwangzihao@163.com); 徐桂琼(1973-), 女, 四川自贡人, 教授, 博士, 主要研究方向为复杂系统建模与动力学研究、数据挖掘、协同过滤。

预测评分 $\hat{R}_{i,j}$, 如式(1)所示:

$$\hat{R}_{i,j} = \sum_{f=1}^k u_{i,f} \cdot v_{j,f} \quad (1)$$

其中: k 表示因子分解的维度, 且 $k \leq \min(m, n)$ 。

在此基础上, Paterek^[13] 引入了代表用户与项目的偏差项来描述评分偏差问题, 如式(2)所示。

$$\hat{R}_{i,j} = \mu + b_i + b_j + \sum_{f=1}^k u_{i,f} \cdot v_{j,f} \quad (2)$$

其中: 参数 b_i 和 b_j 分别代表用户 i 和项目 j 的评分偏差, 常数 μ 表示训练集中的评分值的全局平均数。

为了得到所有用户与项目的偏差以及对应的潜在因子向量, 基于偏差的 MF 需要通过随机梯度下降(SGD)^[8] 等方法优化式(3)所示的目标损失函数 L 。

$$L(u, v, b_k) = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n [(R_{i,j} - \hat{R}_{i,j})^2 + \lambda_1 (\|u_i\|_F^2 + \|v_j\|_F^2) + \lambda_2 (b_i + b_j - \mu)^2] \quad (3)$$

其中后两项是为防止过拟合的正则化项, λ_1 和 λ_2 表示正则化系数, $\|\cdot\|_F^2$ 表示弗罗贝尼乌斯范数。

1.2 因子分解机算法

因子分解机将矩阵因子分解算法拓展为一个通用的预测模型, 即对输入向量 x 的每一组特征进行两两因子分解, 并用分解出来的潜在因子向量的内积和作为预测输出:

$$\hat{y}(x) = \sum_{i=1}^p \sum_{j=i+1}^p x_i x_j \sum_{f=1}^k w_{i,f} \cdot w_{j,f} \quad (4)$$

其中: p 表示输入向量 x 的维数, w_i 表示 x 中第 i 个分量的潜在因子向量。可以看到, 传统矩阵因子分解等算法都是因子分解机在 $p = m + n$ 时的一个子集^[15]。因子分解机的一个简单样例如图 1 所示, 其中 x_1 可以稀疏表示为

$$x_1: \{(\text{feature1}, \text{value1}), (\text{feature2}, \text{value1}), (\text{feature3}, \text{value2})\} \rightarrow y_1$$

由于向量 x 是稀疏的, 式(4)中的大部分项为 0, 这样就快速地对所有特征间交互作用的建模。由文献[14]可知, 式(4)等价于式(5), 可以将其化简至线性的时间复杂度 $O(kp)$ 。

$$\hat{y}(x) = \frac{1}{2} \sum_{f=1}^k [(\sum_{i=1}^p w_{i,f} x_i)^2 - \sum_{i=1}^p w_{i,f}^2 x_i] \quad (5)$$

因子分解机除了可以采用传统的 SGD 等方法优化求解之外, 文献[16]还提出了一种马尔可夫链蒙特卡洛(MCMC)的因子分解机优化方法, 避免了 SGD 中多个参数格搜索的过程, 节省了大量时间。

	X										Y		
	feature1				feature2				feature3				output
	value1	value2	value3	...	value1	value2	value3	...	value1	value2	value3	...	
x_1	1	0	0		1	0	0		0	1	0		y_1
x_2	1	0	0		0	1	0		1	0	0		y_2
x_3	0	1	0		0	1	0		0	1	0		y_3
x_4	0	1	0		0	0	1		0	0	1		y_4
x_5	0	1	0		1	0	0		1	0	0		y_5
x_6	0	0	1		0	0	1		0	0	1		y_6
x_7	0	0	1		1	0	0		0	1	0		
...	...	...	...		...	...	...		...	...	...		...

图1 因子分解机样例

2 基于高阶偏差的因子分解机算法

2.1 基于偏差的用户与项目划分

在实际生活中, 人们对事物的评价因为各自的审美偏好或挑剔程度的不同, 往往存在一种固定的偏好^[11,12], 例如资深电影爱好者就可能具有更为专业且挑剔的眼光, 使得其对电影的评分普遍偏低; 相反普通大众因为不具备专业级的审美水平, 容易给出相对宽松的评价; 也有部分观众较为情绪化, 对喜欢或不喜欢的电影(或其导演、演员)给予极端的评分。同样, 事

物本身的性质或被接受程度也会造成此类偏高、偏低或两极分化的评分倾向性。

在此背景下, 基于偏差的 MF 与其他偏差项算法存在一定的局限性, 一方面通过单一的偏差参数并不能很好地描述现实中用户和项目在评分偏好上的群体特征, 特别是无法区分中等偏好与情绪化用户和争议性项目; 另一方面, 算法中的用户评分偏差与项目评分偏差是相互独立的, 所以求和得出的整体偏差是静态和非个性化的^[14]。因此, 本文提出一种基于偏差的用户与项目划分方法, 根据上述评分偏差的现实特征将用户划分为高评分偏好(high preference, HP)、中评分偏好(medium preference, MP)、低评分偏好(low preference, LP)以及情绪化(emotional, EM)四类; 同样, 将项目划分为高质量(high quality, HQ)、中质量(medium quality, MQ)、低质量(low quality, LQ)和争议(controversial, CV)四类。即用户 u 的偏好类别可以表示为 $P_u \in \{HP, MP, LP, EM\}$, 项目 i 的质量类别可以表示为 $Q_i \in \{HQ, MQ, LQ, CV\}$ 。

基于偏差的用户与项目划分算法描述如下:

输入: 用户集合 $U\{u_1, u_2, \dots, u_m\}$; 项目集合 $I\{i_1, i_2, \dots, i_n\}$; 比例参数 $\alpha \in (0, 0.5)$; 方差阈值 $\beta \in \mathbb{R}$ 。

输出: 用户偏好划分 $P_1, P_2, \dots, P_m$; 项目质量划分 $Q_1, Q_2, \dots, Q_n$ 。

a) 计算用户与项目的评分均值 $\bar{u}, \bar{i}$, 且假设评分均值大小按其编号从小到大排列, 即有 $\bar{u}_1 \leq \bar{u}_2 \leq \dots \leq \bar{u}_m, \bar{i}_1 \leq \bar{i}_2 \leq \dots \leq \bar{i}_n$;

b) 对每个用户 $u_g \in U$

if $g \leq \alpha \times m$, then $P_g \leftarrow LP$;
else if $g \geq (1 - \alpha)m$, then $P_g \leftarrow HP$;

else

计算 u_g 的评分方差 $\sigma_{u_g}^2$;

if $\sigma_{u_g}^2 \leq \beta$, then $P_g \leftarrow MP$;

else $P_g \leftarrow EM$;

c) 对每个项目 $i_h \in I$:

if $h \leq \alpha \times n$, then $Q_h \leftarrow LQ$;

else if $h \geq (1 - \alpha)n$, then $Q_h \leftarrow HQ$;

else

计算 i_h 的评分方差 $\sigma_{i_h}^2$;

if $\sigma_{i_h}^2 \leq \beta$, then $Q_h \leftarrow MQ$;

else $Q_h \leftarrow CV$ 。

2.2 高阶偏差的因子分解机算法

根据 2.1 节中提出的划分方法可以获得用户与项目在评分偏好与本身质量上的偏差类别划分。为了构建不同偏差类别在矩阵分解模型的预测过程中的高阶交互关系, 需要将其作为输入特征与评分记录进行稀疏编码。因此基于高阶偏差的 FM 算法(HOBFM)的输入格式可以表示为

$$x: \{(\text{user}, u), (\text{item}, i), (P, P_u), (Q, Q_i)\} \rightarrow y$$

如图 2 所示的样例中的每一行代表一条评分记录, 对应的 y 代表评分值。以第一行为例, 表示用户 u_1 对项目 i_1 评分为 3 分, 且用户 u_1 在训练集中评分偏好 P_1 为 HP, 项目 i_1 在训练集中的质量 Q_1 表现为 MQ。

	x				P				Q				y
	user	item	P	Q	HP	MP	LP	EM	HQ	MQ	LQ	CV	
x_1	u_1	i_1	3		1	0	0	0	0	1	0	0	3
x_2	u_1	i_2	1		1	0	0	0	0	0	0	1	2
x_3	u_2	i_1	0		0	1	0	0	0	0	0	1	5
x_4	u_2	i_2	0		0	0	1	0	0	0	0	1	1
x_5	u_3	i_1	0		0	0	0	1	0	0	1	0	4
x_6	u_3	i_2	0		0	0	0	0	1	0	0	1	2
x_7	u_4	i_1	0		0	0	0	0	1	0	0	0	3
...	...	...	...	...	...	...	...	...	...	...	...	...	...

图2 HOBFM样例

根据式(4), HOBFM 中用户、项目、用户所属的偏好类别以及项目所属的质量类别之间将进行两两因子分解, 从而得到不同评分偏差的高阶交互关系, 最终的预测评分矩阵 $\hat{R}$ 可以表示为图 3 所示的六组潜在因子矩阵乘积的和。

图3中,(a)即为式(1)所示的基本矩阵因子分解模型;(b)为该用户在其偏好类别中所处的位置因素或归属程度;(c)表示该用户对不同质量类别项目的偏好程度;(d)反映项目在其质量类别中的位置因素或归属程度;(e)表示不同偏好类别的用户群体对该项目的偏好程度;(f)表明不同偏好特征的用户群体对不同偏好特征的项目群体的整体影响。另外,图3(b)与(d)还消除了在划分边界上的不同类别用户与项目的差异。

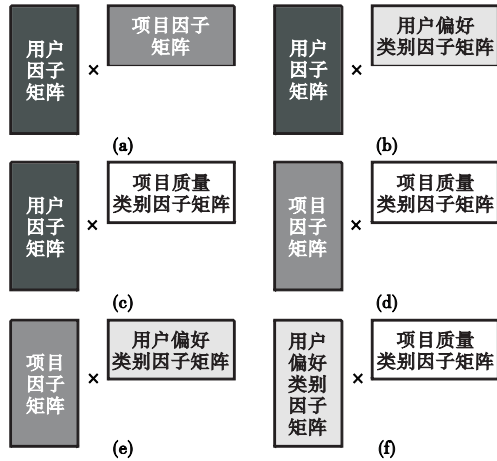


图3 HOBFM中特征的潜在因子交互

3 实验结果与分析

3.1 数据集与评价指标

本文采用 MovieLens 公开数据集中的 ml-100k 和 ml-1m 作为实验数据。其中 ml-100k 数据集包含 943 名用户对 1 684 部电影的 10 000 条评分记录,每名用户至少对 20 部电影评分,每部电影至少被评分 3 次;ml-1m 数据集则包含 6 040 名用户对 3 900 部电影的 1 000 209 条评分记录。数据集中的评分均为 1~5 的整数,数值越大表示用户对该电影的兴趣越高。数据集的稀疏度,即未知评分在整个数据集中的比例,分别为 93.7% 和 95.8%。实验对两个数据集分别进行随机划分,其中 80% 作为训练集,20% 作为测试集。

实验采用均方根误差 (RMSE) 作为评价指标。它通过计算预测的用户评分与实际的用户评分之间的偏差来度量预测的准确性,是推荐算法最常用的准确性指标。指标的值越小,说明预测越准确,算法的性能也就越好。对于测试集 T 中的每一个用户 u 和物品 i ,令 $\hat{R}_{i,j}$ 为算法给出的预测评分, $R_{i,j}$ 为实际评分,则 RMSE 可以通过式(6)计算得出。

$$RMSE = \sqrt{\frac{\sum_{u,i \in T} (R_{i,j} - \hat{R}_{i,j})^2}{|T|}} \quad (6)$$

3.2 HOBFM 的关键参数实验

HOBFM 中的关键参数包含比例参数 α 以及方差阈值 β 。实验采用 ml-100k 数据集,首先固定 β 为数据中用户与项目评分方差的平均值 $\bar{\sigma}^2$,测试不同取值的 α 对 HOBFM 算法 RMSE 的影响。根据图 4 的实验结果可以看出,当比例参数 α 为 0.25 时,HOBFM 算法能达到较好的预测准确度。然后固定 α 为 0.25,测试方差阈值 β 在 $\bar{\sigma}^2$ 的基础上变化对算法 RMSE 的影响。从图 5 的实验结果可以看出,当 β 为 $\bar{\sigma}^2$ 时 HOBFM 算法的准确度较高。

确定参数 α 和 β 后,可利用 2.1 节中给出的算法对用户进行划分,图 6 和 7 分别描述了不同类别用户的评分值的比例成

分,以及不同评分值中的用户类别的比例成分。从图中可以看出经划分后的用户群体已经能明显地体现出与其类别相符合的评分偏差特征。

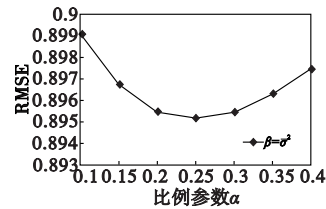


图4 比例参数 α 对HOBFM的影响

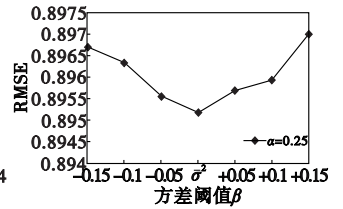


图5 方差阈值 β 对HOBFM的影响

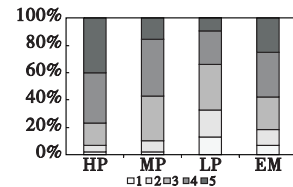


图6 不同类别用户的评分值比例成分图

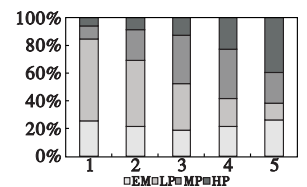


图7 不同评分值的用户类别比例成分图

3.3 HOBFM 算法与经典算法的比较

实验选取基于偏差的 MF (biased MF)、PMF^[7] 以及 SVD + ^[8] 算法作为基准算法,与本文的 HOBFM 算法进行比较。根据 3.2 节的实验结果,设定 HOBFM 算法的比例参数 α 为 0.25、方差阈值 β 为 $\bar{\sigma}^2$,且四个算法的因子分解维度 k 均为 8。

实验的数据集除 ml-100k 和 ml-1m 的所有数据 (all to all) 以外,还分别用 2.1 节中的方法划分为情绪化用户对全体项目 (EM to all)、全体用户对争议项目 (all to VC) 以及情绪化用户对争议项目 (EM to VC) 这三个子集,用以检验各算法针对特殊类别用户、项目的预测性能,实验结果如图 8 和 9 所示。

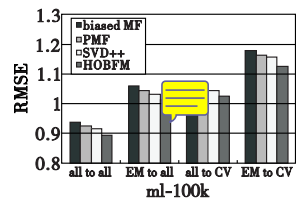


图8 三种算法在ml-100k数据集下的RMSE表现

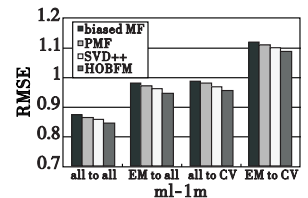


图9 四种算法在ml-1m数据集下的RMSE表现

从图 8 和 9 可以看出,不论是原始数据集还是针对特殊类别群体的子数据集,本文提出的 HOBFM 算法都具有相比 biased MF、PMF 以及 SVD++ 算法更低的 RMSE 指标,说明 HOBFM 算法具有更好的评分预测能力和推荐性能。此外可以看出,这四种算法在预测情绪化用户与争议项目上的准确度均比整体水平要低,但比较而言 HOBFM 算法在其他经典算法的基础上仍有较大的提升。

4 结束语

本文提出的基于高阶偏差的因子分解机推荐算法,首先根据历史评分记录中的评分偏差特征,将用户和项目分别划分为高、中、低评分偏好和情绪化以及高、中、低质量和争议四类。这种划分方法既具有统计学意义,又符合现实中的群体特征;然后将用户与项目的偏差类别作为输入特征集成到因子分解机模型中,建立相比线性模型更高阶的潜在因子模型,改进算法的评分预测。实验结果表明,与其他三种经典矩阵因子分解算法相比,基于高阶偏差的因子分解机算法具有更高的预测准确度,同时在预测情绪化用户与争议项目这类特殊群体上也有一定程度的提升。如何进一步对本文提出的算法进行优化,并

更好地利用历史评分信息、更准确地预测情绪化用户与争议项目,是值得进一步研究的问题。

参考文献:

- [1] 刘建国,周涛,汪秉宏. 个性化推荐系统的研究进展[J]. 自然科学研究进展,2009,19(1):1-15.
- [2] Resnick P, Iacovou N, Suchak M, *et al.* GroupLens: an open architecture for collaborative filtering of netnews[C] //Proc of ACM Conference on Computer Supported Cooperative Work. New York: ACM Press,1994:175-186.
- [3] 查九,李振博,徐桂琼. 基于组合相似度的优化协同过滤算法[J]. 计算机应用与软件,2014,31(12):323-328.
- [4] 李振博,徐桂琼,查九. 基于 Nystrom 扩展谱聚类社会化推荐算法[J]. 计算机应用研究,2015,32(11):3238-3241.
- [5] 胡必云,李舟军,王君,等. 评分偏差对于推荐质量的影响[J]. 北京航空航天大学学报,2012,51(4):823-828.
- [6] Funk S. Netflix update: try this at home[EB/OL]. (2006-12-11). <http://sifter.org/~simon/journal/20061211.html>.
- [7] Mnih A, Salakhutdinov R. Probabilistic matrix factorization[C] //Advances in Neural Information Processing Systems. Cambridge: MIT Press,2007:1257-1264.
- [8] Koren Y. Factorization meets the neighborhood: a multifaceted collaborative filtering model [C] //Proc of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press,2008:426-434.
- [9] 李卫平,杨杰. 基于随机梯度矩阵分解的社会网络推荐算法[J]. 计算机应用研究,2014,31(6):1654-1656.
- [10] 查九,李振博,徐桂琼. 基于用户近邻约束的矩阵因子分解算法[J]. 计算机技术与发展,2015,25(6):1-5.
- [11] Shan Hanhuai, Banerjee A. Generalized probabilistic matrix factorizations for collaborative filtering[C] //Proc of the 10th IEEE International Conference on Data Mining. Washington DC: IEEE Computer Society,2010:1025-1030.
- [12] Park K W, Kim B H, Park T S, *et al.* Uncovering response biases in recommendation[C] //Proc of the 28th AAAI Workshop on Advances in Preference Handling. [S. l.]: AAAI Press,2014:73-78.
- [13] Paterek A. Improving regularized singular value decomposition for collaborative filtering [C] //Proc of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press,2007:5-8.
- [14] Koenigstein N, Dror G, Koren Y. Yahoo! Music recommendations: modeling music ratings with temporal dynamics and item taxonomy [C] //Proc of the 5th ACM Conference on Recommender Systems. New York: ACM Press,2011:165-172.
- [15] Rendle S. Factorization machines[C] //Proc of the 10th IEEE International Conference on Data Mining. Washington DC: IEEE Computer Society,2010:995-1000.
- [16] Rendle S. Factorization machines with libFM [J]. ACM Trans on Intelligent Systems and Technology,2012,3(3):57:1-57:22.
- [17] Rendle S, Gantner Z, Freudenthaler C, *et al.* Fast context-aware recommendations with factorization machines[C] //Proc of the 34th ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press,2011:635-644.
- [18] 余秋宏. 基于因子分解机的社交网络关系推荐研究[D]. 北京:北京邮电大学,2013.
- [19] 胡亚慧,李石君,余伟,等. 一种结合文化和因子分解机的快速评分预测方法[J]. 南京大学学报:自然科学版,2015,51(4):826-833.
- [20] 李振博,徐桂琼,查九. 基于用户近邻约束的矩阵因子分解算法[J]. 计算机技术与发展,2015,25(6):1-5.
- [21] 查九,李振博,徐桂琼. 基于用户近邻约束的矩阵因子分解算法[J]. 计算机技术与发展,2015,25(6):1-5.
- [22] 查九,李振博,徐桂琼. 基于用户近邻约束的矩阵因子分解算法[J]. 计算机技术与发展,2015,25(6):1-5.
- [23] 查九,李振博,徐桂琼. 基于用户近邻约束的矩阵因子分解算法[J]. 计算机技术与发展,2015,25(6):1-5.
- [24] 查九,李振博,徐桂琼. 基于用户近邻约束的矩阵因子分解算法[J]. 计算机技术与发展,2015,25(6):1-5.
- [25] 查九,李振博,徐桂琼. 基于用户近邻约束的矩阵因子分解算法[J]. 计算机技术与发展,2015,25(6):1-5.
- [26] 查九,李振博,徐桂琼. 基于用户近邻约束的矩阵因子分解算法[J]. 计算机技术与发展,2015,25(6):1-5.
- [27] 查九,李振博,徐桂琼. 基于用户近邻约束的矩阵因子分解算法[J]. 计算机技术与发展,2015,25(6):1-5.
- [28] 查九,李振博,徐桂琼. 基于用户近邻约束的矩阵因子分解算法[J]. 计算机技术与发展,2015,25(6):1-5.
- [29] 查九,李振博,徐桂琼. 基于用户近邻约束的矩阵因子分解算法[J]. 计算机技术与发展,2015,25(6):1-5.
- [30] 查九,李振博,徐桂琼. 基于用户近邻约束的矩阵因子分解算法[J]. 计算机技术与发展,2015,25(6):1-5.
- [31] 查九,李振博,徐桂琼. 基于用户近邻约束的矩阵因子分解算法[J]. 计算机技术与发展,2015,25(6):1-5.
- [32] Zhao Yongli, Zhang Yungui, Tong Weiming, *et al.* An improved feature selection algorithm based on MAHALANOBIS distance for network intrusion detection [C] //Proc of International Conference on Sensor Network Security Technology and Privacy Communication System. 2013:69-73.
- [33] Chebrolu S, Abraham A, Thomas J P. Feature deduction and ensemble design of intrusion detection systems[J]. Computers & Security,2005,24(4):295-307.
- [34] Das V, Pathak V, Sharma S, *et al.* Network intrusion detection system based on machine learning algorithms[J]. International Journal of Computer Science & Information Technology, 2010,2(6):138-151.
- [35] McParland C, Peisert S, Scaglione A. Monitoring security of networked control systems: it's the physics[J]. IEEE Security & Privacy,2014,12(6):32-39.
- [36] 赵华. 工业控制系统异常检测算法研究[D]. 北京:冶金自动化研究设计院,2013.
- [37] MacDermott Á, Shi Qi, Merabti M, *et al.* Intrusion detection for critical infrastructure protection[C] //Proc of the 13th Annual Postgraduate Symposium on Convergence of Telecommunications, Networking and Broadcasting. 2012.
- [38] 李骏骁. 多层差异网络深度入侵数据挖掘方法研究[J]. 计算机仿真,2015,32(4):235-238.
- [39] 王海凤. 工业控制网络的异常检测与防御资源分配研究[D]. 杭州:浙江大学,2014.
- [40] 罗耀峰. 面向工业控制系统的入侵检测方法的研究与设计[D]. 杭州:浙江大学,2013.
- [41] Shang Wenli, Zeng Peng, Ming Wan, *et al.* Intrusion detection algorithm based on OCSVM in industrial control system[J]. Security and Communication Networks,2016,9(10):1040-1049.
- [42] 尚文利,张盛山,万明,等. 基于 PSO-SVM 的 Modbus TCP 通讯的异常检测方法[J]. 电子学报,2014,42(11):2314-2320.
- [43] 尚文利,李琳,万明,等. 基于优化单类支持向量机的工业控制系统入侵检测算法[J]. 信息与控制,2015,44(06):678-684.
- [44] Maglaras L A, Jiang J, Cruz T J. Combining ensemble methods and social network metrics for improving accuracy of OCSVM on intrusion detection in SCADA systems[J]. Journal of Information Security and Applications,2016,30(10):15-26.
- [45] Elhag S, Fernández A, Bawakid A, *et al.* On the combination of genetic fuzzy systems and pairwise learning for improving detection rates on intrusion detection systems[J]. Expert Systems with Applications,2015,42(1):193-202.
- [46] Aburomman A A, Reaz M B I. A novel SVM-kNN-PSO ensemble method for intrusion detection system[J]. Applied Soft Computing, 2016,38(1):360-372.
- [47] Singh P, Garg S, Kumar V, *et al.* A testbed for SCADA cyber security and intrusion detection[C] //Proc of International Conference on Cyber Security of Smart Cities, Industrial Control System and Communications. 2015:1-6.
- [48] Rajesh R N D, Selvalakshmi C. NIDS: network intrusion detection system with deceptive virtual hosts for industrial control networks[J]. Scientific Engineering and Applied Science,2015(1):255-263.
- [49] 周晓敏,李璇,黄双. 工业控制系统信息安全仿真平台的设计与实现[J]. 可编程控制器与工厂自动化,2015,21(4):35-40.
- [50] Reaves B, Morris T. An open virtual testbed for industrial control system security research[J]. International Journal of Information Security,2012,11(4):215-229.

(上接第 页)