

引用格式: 靳冠坤, 库涛, 温广波, 等. 隐式反馈场景下基于 Pairwise 排序学习的因子分解机算法[J]. 科学技术与工程, 2018, 18(16): 217—222
Jin Guankun, Ku Tao, Wen Guangbo, et al. Factorization machines based on Pairwise learning with implicit feedback[J]. Science Technology and Engineering, 2018, 18(16): 217—222

隐式反馈场景下基于 Pairwise 排序学习的因子分解机算法

靳冠坤^{1,2} 库涛^{1*} 温广波³ 贾敬崧^{1,2}

(中国科学院沈阳自动化研究所¹ 沈阳 110016; 中国科学院大学² 北京 100049; 渤海造船厂集团有限公司科技与信息部³ 葫芦岛 125004)

摘要 针对隐式反馈场景下的推荐问题以及如何融入用户物品的上下文信息来进行推荐,提出了一种结合 Pairwise 排序学习与因子分解机的隐式反馈推荐模型。首先借鉴排序学习中 Pairwise 的方法解决隐式反馈负反馈缺失的问题,然后选择因子分解机作为排序函数来建模用户的上下文信息,从优化物品排序的角度出发建模用户偏好,进而针对不同用户进行个性化推荐。最终实验也表明,所提出模型在排序指标 MAP 和 NDCG 上都要优于其他 3 种对比算法,在解决隐式反馈下推荐问题的同时,可以利用用户的上下文信息进一步提高推荐的准确度。

关键词 隐式反馈 上下文 个性化推荐 排序学习 因子分解机
中图分类号 TP391.4; **文献标志码** A

随着互联网与电子商务的快速发展,“信息过载”问题日益严重,如何帮助用户快速发现其所需信息,推荐系统应运而生。好的推荐系统对提升用户体验与商品销量等而言都有不可或缺的作用。故推荐系统在工业界得到广泛应用,也是学术界的研究热点之一。

个性化推荐系统通过建模用户偏好,计算用户对推荐物品的评分效用,进而进行推荐^[1]。用户偏好建模的关键资源是用户的历史行为数据,其具体可分为两类:显式反馈(explicit feedback)与隐式反馈(implicit feedback)^[2]。显式反馈特指于那些用户主动提供能准确反映用户偏好的信息,例如物品评分(0~5分)、商品评论等。传统的推荐算法例如协同过滤^[3]、矩阵分解^[4]等,大都是基于用户的显式评分数据来学习用户偏好。与之相对,隐式反馈特指那些隐含表达用户偏好的信息,如商品点击、浏览或购买等行为数据。相比于显式反馈,隐式反馈数据具有收集成本低、数据规模大,更符合实际应用场景的特点。但隐式反馈数据的问题是其往往是单类的(one-class)^[5],即只有用户对物品的正反馈,传统的协同过滤方法无

法很好地适用于该场景,因为隐式反馈场景下用户有过反馈的数据远远小于用户没有过反馈的数据量,从而导致过拟合。针对隐式反馈下的推荐问题,有研究借鉴排序学习^[6](learning to rank)的方法,将推荐任务转化为推荐列表的优化排序问题。Rendle 等^[7,8]最早提出贝叶斯个性化排序模型(Bayesian personalized ranking);该模型通过最大化正例排在负例前面的概率,给出了协同排序的通用框架。Qiang 等^[9]将协同排序成功应用于微博推荐任务中。与前面的方法将推荐转化为 Pairwise 排序学习问题不同,Shi 等^[10]在基于 Listwise 排序学习方法上也进行了研究,提出了 CLiMF 协同过滤推荐算法,该算法直接针对列表的排序指标 MRR 进行优化。

上述基于排序学习的隐式反馈模型大都基于用户物品的二维矩阵来进行推荐,而没有考虑用户与物品交互时的上下文信息,例如时间、地点等。例如用户在寻找餐馆时其与系统交互的位置信息是重要的上下文信息。Adomavicius 和 Tuzhilin 等较早的就指出^[11,12]把上下文信息融入推荐系统将有利于提高推荐的准确度,并提出被广泛引用的“上下文感知推荐系统”的概念。在上下文感知推荐领域中,针对上下文信息建模这一问题,不同的学者提出了各种解决方法,例如上下文预过滤,上下文后过滤,TimeSVD,张量分解模型^[13]等。前两种方法是基于协同过滤的混合方法,其最大的缺点是会造成信息的丢失并需要过多的人工干预操作。后两者为基于模型的方法,其

2017年12月13日收到

国家重点研发计划
(2017YFB030640X)资助

第一作者简介:靳冠坤(1994—),硕士研究生。研究方向:推荐系统。E-mail: jin416789830@gmail.com。

*通信作者简介:库涛(1979—),博士,研究员,博士研究生导师。研究方向:大数据、社会计算、感应网络。E-mail: kutao@sia.cn。

缺点是时间复杂度过高且不利于扩展。目前而言,由 Steff Rendle 提出的因子分解机模型^[14,15];其能够建模各种类型的上下文信息,并具有线性的复杂度。但因子分解机最早是基于用户的显式评分提出的,无法直接应用于隐式反馈推荐场景中。

针对上述问题,本文在隐式反馈的场景下,提出将因子分解机融入到 Pairwise 排序学习方法中,提出了对级学习因子分解机模型(pairwise learning factorization machine, PLFM)。最后在 Movielens-1M 电影数据集上,选用 MAP、NDCG 值作为评价指标,测试了算法的有效性以及与对比实验的差异性。

1 基于 Pairwise 排序学习的因子分解机算法

在介绍对级学习因子分解机模型 PLFM 之前,先给出因子分解机模型的定义与公式,方便后续对本文算法的阐述。

因子分解机(factorization machines)是一种通用的因式分解模型,其借鉴矩阵分解的思想,将每个参数用一个隐向量表示,用隐向量之间的点积表征组合特征的特征参数,可以用于解决分类、回归以及排序问题。采用二阶因子分解机模型,其模型方程如下:

$$\hat{y}(x) = w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle v_i, v_j \rangle < v_i, v_j \rangle > x_i x_j \quad (1)$$

式(1)中,模型参数如下:

$$\begin{cases} w_0 \in R \\ w \in R^n \\ V \in R^{n \times k} \end{cases} \quad (2)$$

式(2)中, n 为特征维度, k 为矩阵 V 分解维度的超参数, $\langle \cdot, \cdot \rangle$ 表示两个 k 维向量之间的点积,表示如下:

$$\langle v_i, v_j \rangle = \sum_{f=1}^k v_{if} v_{jf} \quad (3)$$

另外,式(3)可以通过简单的数学变换使其具有线性的复杂度 $O(kn)$,给出变换后的公式为

$$\hat{y}(x) = w_0 + \sum_{i=1}^n w_i x_i + \frac{1}{2} \sum_{j=1}^k \left[\left(\sum_{i=1}^n v_{if} x_i \right)^2 - \sum_{i=1}^n v_{if}^2 x_i^2 \right] \quad (4)$$

1.1 对级学习因子分解机模型

Pairwise 排序学习的主要思想是将排序问题形式化为二分类问题。通过判断任意两个物品对是否满足偏序关系来进行排序。借鉴文献[6]中 BPRMF 算法中的方法。假设用户对有过反馈的物品集合的偏好大于用户没有过反馈的物品集合的偏好,基于该假设构造物品对作为训练样本。

为更好地描述,做如下定义,令张量矩阵 $S \subset U \times I \times C$ 表示用户的正反馈集合,其中 U 表示用户集合, I 表示物品集合, C 表示上下文信息,例如时间、地点等。又令 I_u^+ 表示用户 u 有过的反馈的物品集合,相反, $I \setminus I_u^+$ 表示相对商品全集而言,用户没有过反馈的物品集合。对于有不同偏好的物品对之间的关系定义为

$$i >_u j \Leftrightarrow i \in I_u^+ \cap j \in I \setminus I_u^+ \quad (5)$$

根据上述假设,可以从数据集中提取对级偏好对,定义为

$$(u, i, j) \in S \Leftrightarrow i \in I_u^+ \cap j \in I \setminus I_u^+ \quad (6)$$

在定义损失函数之前先定义两个概率值:预测排序概率 P_{ij} 与真实排序概率 \bar{P}_{ij} 。令物品对的预测后验概率:

$$P_{ij} = P(x^i >_u x^j) = 1 / \{ 1 + \exp[-(s_i - s_j)] \} \quad (7)$$

式(7)中 s_i, s_j 为模型预测分数。令 \bar{P}_{ij} 为已知的排序概率,即相对于物品 j ,用户 u 对物品 i 有更强的偏好的概率值。令 $s_{ij} = s_i - s_j$,则有如下关系:

$$\bar{P}_{ij} = \frac{1}{2} (1 + s_{ij}) \quad (8)$$

如果用户 u 更喜欢物品 i ,则 $s_{ij} = 1$;如果用户更喜欢物品 j 则 $s_{ij} = -1$ 。另有输入向量 $x \in R^n$,其中 n 表示特征数量。输出空间 $y \in \{-1, +1\}$,其中 $+1$ 表征用户对 i 的偏好大于 j , -1 表示相反的情况。

在损失函数的选取上,借鉴了 RankNet^[16]中的方法,该方法的基本假设是,对于一个推荐列表,排序结果越好,那么有错误相对关系的物品对就越少,以最小化错误物品对的数量作为优化目标。对于每一个物品对,使用交叉熵 C 来度量其预测代价,定义如下:

$$C_{ij} = -\bar{P}_{ij} \lg P_{ij} - (1 - \bar{P}_{ij}) \lg (1 - P_{ij}) \quad (9)$$

式(9)中, C_{ij} 表示预测概率 P_{ij} 对真实排序概率 \bar{P}_{ij} 的拟合程度。

将式(7)、式(8)代入式(9)化简可得:

$$C_{ij} = \frac{1}{2} (1 - s_{ij}) (s_i - s_j) + \lg [1 + e^{-(s_i - s_j)}] \quad (10)$$

与 RankNet 和 BPRMF 等算法不同的是,RankNet 中排序函数选择了神经网络,其不适用于推荐场景下数据稀疏性问题。而 BPRMF 中其使用矩阵分解模型作为排序函数,无法利用各种上下文信息。而这里选择因子分解机作为排序函数,来建模上下文信息。

由式(10)可知,目标函数只用到物品对之间的模型分数之间的差值,将因子分解机作为排序函数,将式(4)代入计算可得 s_{ij} :

$$s_{ij} = \sum_{i=1}^n w_i [x_i^{(i)} - x_i^{(j)}] - \frac{1}{2} \sum_{f=1}^k \left[\sum_{i=1}^n v_{i,f}^2 x_i^{(i)2} - \sum_{i=1}^n v_{i,f}^2 x_i^{(j)2} \right] + \frac{1}{2} \sum_{f=1}^k \left\{ \left[\sum_{i=1}^n v_{i,f} x_i^{(i)} \right]^2 - \left[\sum_{i=1}^n v_{i,f} x_i^{(j)} \right]^2 \right\} \quad (11)$$

式(11)中,下标 i, j 表示样本 x 的第 i, j 个特征;上标 i, j 表示样本对中第 i, j 个样本。

由于分解模型的模型参数过多容易产生过拟合现象,故在损失函数中为每个参数引入一个正则项 λ_θ ,再将式(11)代入交叉熵损失函数式(10)中,可得最终要优化求解的目标函数:

$$C = \sum_{i,j \in S} C_{ij} + \sum_{\theta \in \Theta} \lambda_\theta \|\theta\|^2 \quad (12)$$

式(12)中, S 为训练物品对集合, $\Theta = \{w_0, w_1, \dots, w_n, v_{1,1}, \dots, v_{n,k}\}$ 为模型参数, $\|\cdot\|^2$ 为参数矩阵的 2-范数, λ_θ 为 L2 正则项的超参数。

1.2 模型参数的优化求解

使用随机梯度下降法(SGD)来优化上述损失函数求解模型参数。对式(14)求导可得参数更新公式为

$$\theta \rightarrow \theta - \eta \left(\frac{\partial C_{ij}}{\partial \theta} + \lambda_\theta \theta \right) \quad (13)$$

式(13)中:

$$\frac{\partial C_{ij}}{\partial \theta} = \frac{\partial C_{ij}}{\partial s_i} \frac{\partial s_i}{\partial \theta} + \frac{\partial C_{ij}}{\partial s_j} \frac{\partial s_j}{\partial \theta} \quad (14)$$

又由式(10)可得 $\frac{\partial C_{ij}}{\partial s_i}$ 与 $\frac{\partial C_{ij}}{\partial s_j}$ 为

$$\frac{\partial C_{ij}}{\partial s_i} = \left(\frac{1 - s_{ij}}{2} - \frac{1}{1 + e^{s_i - s_j}} \right) = - \frac{\partial C_{ij}}{\partial s_j} \quad (15)$$

由式(13)~式(15)可得:

$$\theta \rightarrow \theta - \eta \left[\left(\frac{1 - s_{ij}}{2} - \frac{1}{1 + e^{s_i - s_j}} \right) \frac{\partial (s_i - s_j)}{\partial \theta} + \lambda_\theta \theta \right] \quad (16)$$

又易得因子分解机关于各参数的梯度为

$$\frac{\partial \hat{y}(x)}{\partial \theta} = \begin{cases} 1, & \theta = w_0 \\ x_i, & \theta = w_i \\ x_i \sum_{j=1}^n x_j v_{j,f} - v_{i,f} x_i^2, & \theta = v_{i,f} \end{cases} \quad (17)$$

合并式(16)~式(17)可得参数更新方程如下:

$$w_i \rightarrow w_i - \eta \left[\left(\frac{1 - s_{ij}}{2} - \frac{1}{1 + e^{s_i - s_j}} \right) (x_i^{(i)} - x_i^{(j)}) + \lambda_{w_i} w_i \right] \quad (18)$$

$$v_{i,f} \rightarrow v_{i,f} - \eta \left\{ \left(\frac{1 - s_{ij}}{2} - \frac{1}{1 + e^{s_i - s_j}} \right) \times \left[\sum_{j=1}^n v_{j,f} (x_i^{(i)} x_j^{(i)} - x_i^{(j)} x_j^{(j)}) - v_{i,f} (x_i^{(i)} - x_i^{(j)}) \right] + \lambda_{v_{i,f}} v_{i,f} \right\} \quad (19)$$

1.3 PLFM 算法流程

由前所述,使用随机梯度下降法求解 PLFM 模型的参数,其基本算法流程如下:

算法 1 SGD 求解 PLFM

- (1) 输入: 训练数据集、正则化参数 λ 、学习速率 η 以及参数 σ 。
- (2) 输出: FM 模型的参数 $\theta = (w, V)$ 。
- (3) 初始化参数 $w = (0, 0, \dots, 0)$, $V \sim N(0, \sigma)$ 。
- (4) 从训练集中依次选取训练样本对 (u, i) 、 (u, j) 。
- (5) 计算 $s_i = \hat{y}[x^{(i)}]$, $s_j = \hat{y}[x^{(j)}]$ 。
- (6) 由式(18)和式(19)更新参数 w, V 。
- (7) 重复步骤(4)~步骤(6)直至收敛。

上述算法的时间复杂度为 $O(|U| |I^+| |I \setminus I^+|)$ 其中 I^+ 为用户 $u \in U$ 有过反馈的物品数量, $I \setminus I^+$ 为用户 u 相对于商品全集 I 中没有过反馈的物品数量,显然这是一个非常大的数字,会导致训练速度过慢。为解决这个问题,针对算法 1 中的步骤(4),设计了一种采样方案来解决这一问题。具体方案如下。

算法 2 训练样本的随机采样方案

- (1) 从用户集合 U 中采样一个用户 u 。
- (1) 从用户有过反馈的物品集合 I^+ 中随机采样出物品 i 。
- (3) 从用户没有过反馈的物品集合 $I \setminus I^+$ 随机采样出物品 j 。

故在实际的参数求解中,将算法 1 的步骤(4)替换为算法 2,即为最终的算法流程。

2 实验与结果

2.1 数据集

为验证本文算法在隐式反馈场景下的有效性,以及在模型中融入上下文信息可以提高准确性,同时也为了便于与现有方法进行对比,采用了推荐领域常用的数据集—Movielens—1 M 电影数据集。该数据集包含 6 040 个用户和 3 706 部电影,共 1 000 209 条评价记录,电影评分为 1~5。数据集中包含用户的内容信息(年龄、性别、职业),电影内容信息(电影类型、观看时间、地点)等作为上下文信息。另外为把该数据集转换成隐式反馈数据集,借鉴文献[17]中的做法,将用户有过评分的数据值设为 1,表征用户对其有过反馈,没有评分的数据值设为 0,表征用户对其没有过反馈。所有数据均采用 one-hot-encoding 编码方式,数据划分 80% 作为训练集,20% 作为测试集。

2.2 对比实验的选择

为验证提出算法的有效性,使用了其他三个算法作为对比实验:MP(most popular)、FM、BPRMF。其中 MP 为向用户推荐热门物品,作为 Baseline。FM 为因子分解机模型,与提出的 PLFM 作为对比,来对比 Pairwise 方法的有效性。BPRMF 为 Steff Rendle 提出的基于 Pairwise 方法的

矩阵分解模型,将其作为对比实验来判断在模型中融入上下文信息对提高推荐准确度的影响。

2.3 评价指标

提出的算法都是基于提高推荐列表的排序质量,因为对于推荐结果而言,用户更看重推荐列表中排序靠前的物品的推荐质量,故选取了用于评价排序质量的两个常用指标 MAP 和 NDCG。

MAP(mean average precision) 是平均准确率均值。具体计算公式如下:

$$MAP = \frac{1}{r} \sum_{i=1}^r \frac{1}{\text{第 } i \text{ 个相关物品的位置}} \quad (20)$$

式(20)中 r 表示列表中相关物品的总数。

NDCG(normalized discounted cumulative gain) 是归一化折损累积增益。对于给定用户 u ,NDCG 在 N 位的值为

$$NDCG@N = \frac{1}{Z_N} \sum_{i=1}^N \frac{2^{r(i)} - 1}{\lg(1 + i)}$$

式(21)中 N 为列表中的位置, Z_N 为归一化因子, $r(i)$ 为第 i 个位置的物品的相关度。NDCG 与 MAP 的值的取值范围均为 $[0, 1]$,若取值越大,则推荐效果越好。

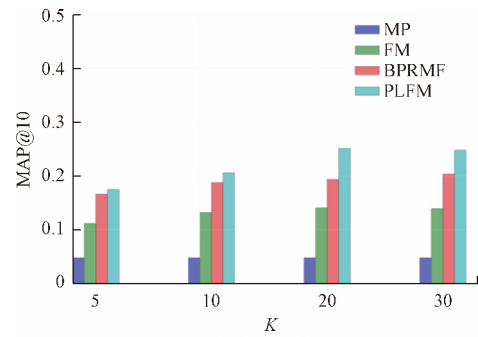
2.4 实验结果分析

2.4.1 性能比较

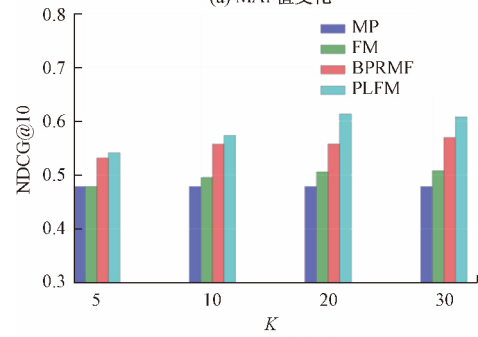
图1展示了对比算法与提出算法在 MAP、NDCG 两种评价指标上的结果。由于除 MP 外,其余三种算法都是矩阵分解模型,故在图1中展示了不同的 k 值下各指标的变化。由结果可知,随着 k 值得增加,FM、BPRMF、PLFM 三种算法在指标上都有所上升,且在 $k = 20$ 时性能达到最优。首先,可以看出提出的 PLFM 算法在两个指标上都要优于原始的 FM 算法,故可以证明基于 Pairwise 方法的因子分解机模型可以更好地解决隐式反馈场景下的推荐问题。另外,相比于未包含上下文信息的 BPRMF 模型,提出的 PLFM 算法在各排序指标上都要更好,也说明了在模型中融入上下文信息对提高推荐准确性的重要性。

2.4.2 收敛性分析

图2比较了在 $k = 20$ 的条件下,各算法随着迭代次数的增加,三个评价指标的变化情况。从图2中可以看出,提出的模型 PLFM 随着迭代次数的增



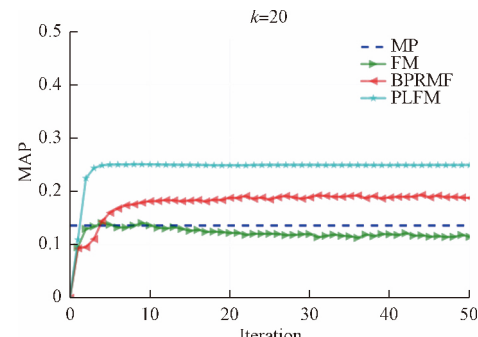
(a) MAP值变化



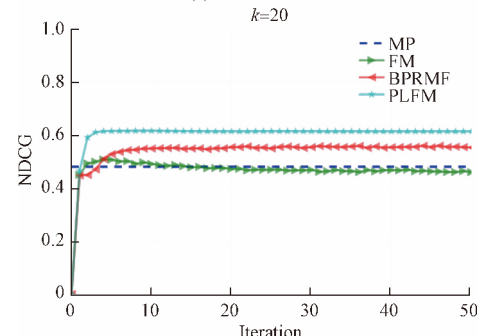
(b) NDCG值变化

图1 各算法在不同评价指标 MAP、NDCG 上以及不同 k 值(5、10、20、30)下的变化

Fig. 1 Performance comparison of each algorithm in different evaluation index MAP, NDCG with different k (5, 10, 20, 30) value



(a) MAP值收敛情况



(b) NDCG值收敛情况

图2 各算法随迭代次数的收敛情况。

Fig. 2 The convergence of each algorithm with the increment of iterations

加,各指标都快速增加并趋于稳定,同时与其他矩阵

分解算法相比而言,也有着较快的收敛速度。

2.4.3 排序质量比较

表 1 展示了不同维度下, $Top-N$ ($N=3, 5, 10, 20$) 推荐的排序结果, 衡量指标为 NDCG 值。可以看出, 相比于次优的 BPRMF 算法, PLFM 算法在模

型中融入了用户物品的内容信息与上下文信息, 性能相比而言提升了约 7% ~ 13%。相比 FM 而言, PLFM 使用 Pairwise 方法, 从优化排序角度的出发来处理隐式反馈场景下负反馈缺失的问题, 性能相比而言提升了约 15% ~ 20%。

表 1 各算法在不同 k 值下的 $Top-N$ ($N=3, 5, 10, 20$) 排序质量

Table 1 The ranking performance comparison of $Top-N$ ($N=3, 5, 10, 20$) with different k value

算法	NDCG@3				NDCG@5				NDCG@10				NDCG@20			
	$k=5$	$k=10$	$k=20$	$k=30$	$k=5$	$k=10$	$k=20$	$k=30$	$k=5$	$k=10$	$k=20$	$k=30$	$k=5$	$k=10$	$k=20$	$k=30$
MP	0.053	0.063	0.099	0.127												
FM	0.186	0.227	0.214	0.222	0.219	0.243	0.262	0.226	0.271	0.258	0.306	0.284	0.303	0.316	0.342	0.326
BPRMF	0.399	0.426	0.436	0.389	0.408	0.453	0.449	0.444	0.421	0.456	0.475	0.463	0.425	0.455	0.461	0.435
PLFM	0.436	0.458	0.490	0.505	0.522	0.433	0.513	0.520	0.460	0.506	0.471	0.484	0.451	0.475	0.504	0.533

3 结论

针对隐式反馈场景下的推荐问题, 在前人工作的基础上提出了基于 Pairwise 排序学习的因子分解机算法。在解决隐式反馈场景下的推荐问题的同时, 使用因子分解机来建模额外的用户物品的内容信息以及上下文信息, 以提供更准确的推荐。并通过实验将该算法与其余三种算法进行对比, 同时在评价准则上选择了更符合实际场景的排序指标, 也都取得了相比对比实验而言更好的结果, 验证了算法的可行性与有效性。

未来工作将就提出的算法做进一步的改进, 现有可改进的点包括有: ①采样方案的设计, 本文在负样本选择上采取了简单的随机抽样, 后续将设计更为合理有效的采样方案, 以加快模型的收敛速度。②排序模型的选择, 本文采用因子分解机作为排序函数, 在原有框架不变的情况下, 依然可以尝试更多其他的模型, 例如神经网络模型, 未来将尝试将深度学习技术应用到隐式反馈场景下的可行性。

参 考 文 献

- Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 2005; 17(6): 734—749
- Gai L, Qiang C. Exploiting explicit and implicit feedback for personalized ranking. *Mathematical Problems in Engineering*, 2016; 2535329: 1—11
- Koren Y. Collaborative filtering with temporal dynamics. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Paris: ACM, 2009: 447—456
- Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. *IEEE Computer Journal*, 2009; 42(8): 30—37
- Pan R, Zhou Y, Cao B, et al. One-class collaborative filtering. *Proceedings of the 8th IEEE International Conference on Data Mining*.

- Pisa: ICDM, 2008: 502—511
- Trotman A. Learning to rank. *Information Retrieval Journal*, 2005; 8(3): 359—381
- Rendle S, Freudenthaler C, Gantner Z, et al. BPR: Bayesian personalized ranking from implicit feedback. *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence*. Virginia: AUAI Press, 2009: 452—461
- Pan W, Chen L. GBPR: Group preference based Bayesian personalized ranking for one-class collaborative filtering. *Proceedings of 23th International Joint Conference on Artificial Intelligence*. Beijing: AAAI, 2013: 2691—2697
- Qiang R, Liang F, Yang J. Exploiting ranking factorization machines for microblog retrieval. *Proceedings of the 22nd ACM International Conference on Information and Knowledge Management*. San Francisco: ACM, 2013: 1783—1788
- Shi Y, Karatzoglou A, Baltrunas L, et al. CLiMF: Learning to maximize reciprocal rank with collaborative less-is-more filtering. *Proceedings of the 6th Conference on Recommender Systems*. New York: ACM, 2012: 139—146
- Adomavicius G, Tuzhilin A. *Recommender systems handbook*. Berlin: Springer-Verlag press, 2011: 217—253
- Adomavicius G, Alexander T, Sen S, et al. Incorporating contextual information in recommender systems using a multidimensional approach. *ACM Transactions on Information Systems*, 2005; 23(1): 103—145
- Karatzoglou A, Amatriain X, Baltrunas L, et al. Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering. *Proceeding of Recsys 2010*. New York: ACM, 2010: 79—86
- Rendle S. Factorization machines. *Proceedings of 13th International Conference on Data Mining*. Sydney: ICDM, 2010: 995—1000
- Rendle S. Factorization machines with libFM. *ACM Transactions on Intelligent Systems and Technology*, 2012; 3(3): 1—22
- Burges C, Shaked T, Renshaw E, et al. Learning to rank using gradient descent. *Proceedings of the 22nd International Conference on Machine Learning*. New York: ACM, 2005: 89—96
- Li Y, Hu Y, Chen Y, et al. Improving one-class collaborative filtering by incorporating rich user information. *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*. Toronto: ACM, 2010: 959—968

Factorization Machines Based on Pairwise Learning with Implicit Feedback

JIN Guan-kun^{1 2}, KU Tao^{1*}, WEN Guang-bo³, JIA Jing-song^{1 2}

(Shenyang Institute of Automation, Chines Academy of Science¹, Shenyang 110016, China;

University of Chines Academy of Science², Beijing 100049, China;

BoHai Shipyard Group Co., Ltd. Ministry of Science and Technology³, Huludao 125004, China)

[Abstract] In view of the problem of personalized recommendation for implicit feedback and how to incorporate users' contextual information in recommendation, a recommendation model combined with pairwise learning and factorization machine was proposed. First of all, the method of pairwise learning was used to solve the problem of negative feedback missing under implicit feedback scenario, and then choose the factorization machine as ranking function to model the user's contextual information. Specifically, the user's preference was modeled from the perspective of optimizing the ranking of items, eventually, provided personalized recommendations for different users according to the model score. Experiments also show that the model proposed is better than the other three contrast algorithms in terms of ranking indicators such as MAP and NDCG. While addressing the recommended problems under implicit feedback, the proposed model can further improve the accuracy of recommendation by using the user's contextual information.

[Key words] implicit feedback context personalized recommendation pairwise learning factorization machines