

加入用户评分偏置的推荐系统排名模型

彭飞^{1,2}, 邓浩江^{1,2}, 刘磊^{1,2}

(1. 中国科学院国家网络新媒体工程技术研究中心, 100190, 北京; 2. 中国科学院声学研究所, 100190, 北京)

摘要: 针对现有推荐技术忽视了推荐结果总体多样性的问题, 提出一种加入用户评分偏置的推荐系统排名模型. 该模型通过交叉验证获取偏置调节参数, 用以计算用户评分偏置, 利用该偏置改进了传统模型中的用户阈值计算方法, 从而保证总体多样性的提升效果可作用于全局用户. 通过在阈值计算环节引入多样性调节因子, 并结合多种排名算法, 实现了总体多样性和精准度的精细控制. 实验结果表明, 相对于传统的基于经验取值的排名模型, 所提模型在保证推荐精准度的同时, 提高了推荐结果的总体多样性, 可以方便地嵌入互动电视和电子商务等推荐系统, 无需对原功能模块进行改动.

关键词: 推荐系统; 精准度; 多样性; 排名模型; 用户评分偏置

中图分类号: TP393 **文献标志码:** A **文章编号:** 0253-987X(2012)06-0074-05

A Rank Model with User Bias for Recommender System

PENG Fei^{1,2}, DENG Haojiang^{1,2}, LIU Lei^{1,2}

(1. National Network New Media Engineering Research Center, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China; 2. Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: A new rank model that takes user's bias into account is proposed to solve the problem that existing rank techniques do not consider the aggregate diversity of the recommendation result. The bias regulating parameter is obtained by cross validation, and is used to calculate the user's bias. Then the algorithm to calculate the user's threshold in the traditional rank models is improved by using the resulting user's bias, so that the enhancement of the aggregate diversity can be effective to all users. The goal of fine control on the aggregate diversity and accuracy is achieved by taking the diversity regulating parameter into the process of calculating the threshold and combining with multiple rank methods. Experimental results and comparisons with the traditional rank model based on experience show that the proposed model can generate recommendations that have higher aggregate diversity across all users while maintaining the recommendation accuracy. Besides, the proposed model can be embedded into interactive TV or E-commerce recommender system without changing the original functional components.

Keywords: recommender system; accuracy; diversity; rank model; user bias

在信息过载的时代, 无论是作为消费者的普通用户, 还是作为信息生产者的内容提供商都遇到了很大的挑战. 推荐系统的任务就是联系用户和信息, 一方面帮助用户发现对自己有价值的信息, 另一方

面让信息能够展现在对它有兴趣的用户面前, 从而实现信息消费者和信息生产者的双赢. 以往对于推荐系统的研究^[1]大都集中于如何提高预测结果的精准度, 然而这样的推荐结果往往比较片面. 通过挖掘

推荐结果其他方面的因素^[2](如多样性、可信度、新颖度等)可以有效地改善推荐结果质量,进而为用户带来更好的体验。本文提出一种改进的排名模型,通过在推荐系统的排名环节中引入用户评分偏置,在保证系统推荐精准度的同时,大幅度提高推荐结果的总体多样性,进而提升系统的推荐质量。

在获取总体多样性提升的同时,本文提出的排名模型还具有两方面的特点:①参数化,该模型可以通过调节模型参数来获取所需的精准度和多样性;②适应性,该模型可方便地嵌入任何推荐系统中,无需对原有系统做出改动。

1 研究背景

1.1 推荐系统概述

根据推荐方法的不同,推荐系统通常分为3类:基于内容的推荐、协同过滤推荐和混合推荐^[3-4]。基于内容的推荐是根据用户过去的行为推断用户的偏好。协同过滤推荐是根据具有相似喜好的其他用户的行为,推断这一类用户的使用偏好。混合推荐是综合使用内容推荐和协同推荐两种方式。

一个推荐系统需要完成两方面的任务^[5]:一方面,需要根据用户的历史记录,利用推荐算法,预测用户对尚未使用过的物品的喜好程度,这是一个评分预测的过程;另一方面,需要利用前一方面对用户的评分预测结果生成最终的推荐列表,在这一过程中,需要综合考虑多方面的因素,如系统的多样性、新颖性、可信度等。本文提出的排名模型正是通过提升推荐结果的多样性来提高系统的推荐质量。

为了对推荐结果做出有效的评估,需要对推荐结果的精准度、多样性等采用量化的计算方法。下面简要介绍本文采用的评估标准。

1.2 推荐精准度

对于推荐系统精准度的评估可以分为两个层次。从评分预测的角度,包括平均绝对误差(MAE)、均方误差(RMSE);在评分预测的基础上,需要完成最终的 top- N 推荐,这方面的评估包括精准度(precision)、召回率(recall)等^[6]。本文采用 top- N 推荐评价指标中的精准度作为评测标准。

1.3 推荐多样性

推荐系统的目的之一就是向用户推荐独特、个性化的产品^[5],而只有多样化的推荐结果才能达到这样的效果。许多研究^[5,7-11]都强调了推荐结果多样性的重要性。推荐结果的多样性可以从个体和总体两个角度进行度量。

个体多样性是指避免向一个用户推荐过于相似的物品。一些研究^[7-9]利用推荐列表的内部相似度来计算个体多样性,文献^[10]采用物品新颖度来计算一个物品对一个推荐列表带来的多样性增量。

然而,推荐结果的个体多样性并不一定能带来总体多样性的提升^[5]。总体多样性是指推荐系统向所有用户的推荐结果的覆盖情况,高的总体多样性不仅能给用户带来更丰富的推荐体验(长尾效应^[12]),同时对于商家来说,更加多样的推荐结果也将带来更加合理的经营模式^[13]。

文献^[14]从用户和产业的角度讨论了总体多样性对产业效果的影响,传统推荐系统由于推荐方式的限制而使得新推出的产品得不到有效的展现,推荐的更多的是长期存在并相对受到好评的物品。文献^[5]通过改变标准排名算法,使系统尽量推荐受到“冷落”的物品,从而提高系统的总体多样性,但未考虑用户评分的差异性对排名模型的影响。本文提出的排名模型在文献^[5]的基础上,通过引入用户评分偏置,在保证精准度的同时,大幅度提高了推荐结果的总体多样性。

2 加入用户评分偏置的排名模型

在推荐系统的应用过程中,为了得到系统推荐多样性的提升,必然带来精准度的下降^[5],推荐系统的各个环节都会对评估指标产生影响。我们针对推荐系统的排名环节,引入一种新的排名模型,在保证推荐精准度的同时,大幅度提升推荐结果的总体多样性。该模型通过参数调节可以获取所需的精准度与多样性。此外,该模型可以方便地应用于电子商务推荐系统以及其他类型的推荐系统,无需对原有系统进行更改。相对于文献^[5]中的排名模型,本文提出的排名模型在性能上有明显的提升。

2.1 模型描述

首先,我们对文中将要使用的符号做出说明。

U 表示推荐系统的用户集合, I 表示被推荐的物品集合。用户评分数据集 $D = \{(u, i, t, r)\}$ 是4元组的集合,其中每一条评分记录,即一个4元组 (u, i, t, r) ,代表用户 $u \in U$ 在 t 时刻给物品 $i \in I$ 赋予评分 r 。 r_{ui} 表示用户 u 对物品 i 的评分, $\hat{r}_{ui}$ 为预测算法对用户 u 给物品 i 评分的预测值。

标准排名算法的思路是根据物品的预测评分按降序排列,即预测评分越高排名越靠前,其排名公式可归纳为

$$k_{\text{Std}}(i) = \hat{r}_{ui}^{-1} \quad (1)$$

即物品的排名与该物品的预测得分成反比。

Adomavicius 等^[5]对式(1)所示的标准排名算法进行了优化,用以提高推荐结果的总体多样性,其基本思路是向用户推荐“冷门”的物品,从而获取系统总体多样性的提升。例如,以物品流行度升序排名的公式为

$$k_{\text{ItemProp}(i)} = |U(i)|$$

$$\text{where } U(i) = \{u \in U \mid \exists r_{ui} \in D\} \quad (2)$$

用评分数据集 D 中对物品 i 有过评分记录的所有用户的数目来表示 i 的流行程度,并按此流行度升序排名,从而将“不流行”的物品排在前面。在此基础上,为了保证推荐结果的精准度,对排名模型进一步优化为

$$k_x(i, T_R) = \begin{cases} k_x(i), & \text{if } \hat{r}_{ui} \in [T_R, T_{\max}] \\ \alpha_u + k_{\text{Std}}(i), & \text{if } \hat{r}_{ui} \in [T_H, T_R) \end{cases}$$

$$\text{where } \alpha_u = \max_{i \in I_u(T_R)} k_x(i)$$

$$\hat{I}_u(T_R) = \{i \in I \mid \hat{r}_{ui} \geq T_R\} \quad (3)$$

式中: $k_x(i)$ 表示用来提高推荐多样性的排名算法。式(2)为物品流行度升序排名算法,其他有效的排名算法还包括^[5]按预测评分升序排名、按物品平均分升序排名、按物品的绝对或相对受欢迎程度升序排名、按物品的评分方差升序排名、按用户近邻物品的评分方差升序排名等。

当用户的预测评分 $\hat{r}_{ui}$ 高于阈值 T_R 时,用 $k_x(i)$ 来计算物品排名, $\hat{r}_{ui}$ 低于 T_R 时,用标准的排名方法 $k_{\text{Std}}(i)$ 来计算物品排名,通过加上 α_u ,即预测评分大于阈值 T_R 的所有物品的排名最大值,保证标准排名算法不影响整体排名趋势。 $T_{\max}$ 是评分系统的最高分,例如在一个 5 分推荐系统中, $T_{\max} = 5$, T_H 是文中提出的用以判定一个物品是否与用户相关的阈值,取经验值 3.5,用户评分大于 T_H ,则认为用户对该物品感兴趣。

上述模型有效地保证了推荐系统在不过分降低推荐精准度的同时提高推荐结果的多样性,文中也给出了相关的测试说明。但是,该排名算法的问题在于,对于所有用户的推荐采用统一的“喜好”评判标准,没有考虑每个用户个体之间评分的差异性。

例如,某用户 u_a 对物品打分比较苛刻,对于他最喜欢的物品也只给 3 分,那么对该用户的预测打分也都会低于 3 分,即大多数 $\hat{r}_{u_a i} < T_H$,式(3)将完全失去效用,而另一个用户 u_b 对于物品打分总是很高,对于最不喜欢的物品打分也会高于 4 分,那么 $\hat{r}_{u_b i} > T_R$ 将总是成立,式(3)将退化为完全按 $k_x(i)$

进行排名,系统的推荐精准度将无法保证。

因此,在排名模型中,必须充分考虑用户打分的差异性影响,保证模型的多样性提升效果可作用于全局用户。为此,我们通过加入用户评分偏置,对式(3)进行改进后得到

$$k_x(i, u) = \begin{cases} k_x(i), & \text{if } \hat{r}_{ui} \in [T_b(\beta), T_{\max}] \\ \alpha_u + k_{\text{Std}}(i), & \text{if } \hat{r}_{ui} \in [0, T_b(\beta)) \end{cases}$$

$$\text{where } T_b(\beta) = (1 - \beta)b_u + \beta T_{\max}$$

$$\alpha_u = \max_{i \in \hat{I}_u(\beta)} k_x(i)$$

$$\hat{I}_u(\beta) = \{i \in I \mid \hat{r}_{ui} \geq T_b(\beta)\} \quad (4)$$

式中: b_u 表示用户 u 的评分偏置值; $T_b(\beta) = (1 - \beta)b_u + \beta T_{\max} = b_u + \beta(T_{\max} - b_u)$ 是对式(3)中 T_R 的演化。对于一个用户 u ,首先根据其历史打分记录计算出他的评分偏置 b_u ,我们认为,只有某物品的预测评分高于 b_u 才表明该用户“喜欢”该物品,通过设置阈值 $T_b(\beta) = (1 - \beta)b_u + \beta T_{\max}$ 来控制最终推荐 top- N 结果的精准度与多样性。 β 为多样性调节因子,表示阈值在 $[b_u, T_{\max}]$ 区间内高出 b_u 的比例,取值范围为 $\left[1 - T_{\max}/(T_{\max} - \min(b_u)), 1\right]$,通过 β 值的变化,可以控制 $T_b(\beta)$ 在 0 到 $T_{\max}$ 间变化,相对于传统的 T_H 到 $T_{\max}$ 的 $\hat{r}_{ui}$ 取值空间,式(4)更为有效地提升了精准度与多样性的可调范围。

评分偏置 b_u 的计算公式为

$$b_u = \frac{\sum_{(u,i) \in D} r_{ui}}{\lambda + |\{i \mid (u,i) \in D\}|} \quad (5)$$

获取 b_u 的方法就是计算用户以往评分记录的平均值,并通过偏置调节参数 λ 来根据应用场景调节 b_u 的大小, λ 可以通过交叉验证得到,在后文测试中取 $\lambda = -4$ 。

2.2 实现步骤

上文描述了基于用户评分偏置的排名模型的基本原理,下面将介绍如何将上述排名模型应用到实际的推荐系统中,实现步骤如图 1 所示。

以向用户 u_2 推荐 top-5 列表为例。首先,系统根据原始评分矩阵 $\mathbf{M}_1$,利用评分预测算法对 u_2 未评分的物品进行预测评分,得到评分矩阵 $\mathbf{M}_2$;接着,从 u_2 的预测评分记录中抽取出尚未评分并且预测评分超过 b_{u_2} 的预测记录。标准排名算法此时就会对评分物品从高到低进行排名,得到 top- N 列表,而优化后的排名模型采取的策略是:首先将排名按照物品流行度(或物品平均分等方式)由低到高排名,即把受“冷落”的物品排在前面,进一步地,把这个排

名列表中预测评分低于阈值 $T_b(\beta) = (1 - \beta)b_u + \beta T_{\max}$ 的物品按照排名公式计算出的排名结果进行替换,最终得到既符合用户“喜好”,又“多样”的 top-N 推荐列表。

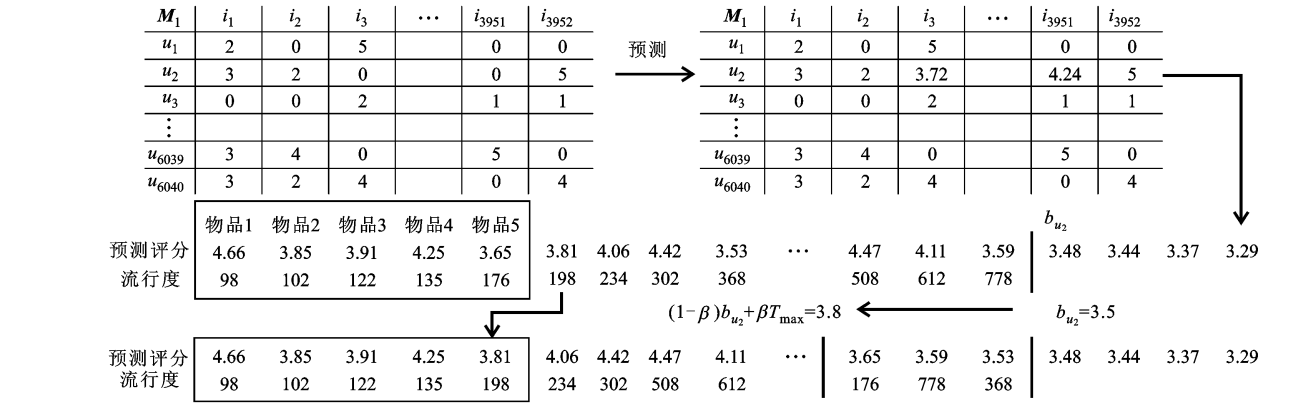


图 1 排名模型实现步骤图

2.3 性能测试

为了对本文描述的排名模型的推荐效果做出有效地评估,我们采用 Movielens 的 1 M 数据集(<http://www.grouplens.com/>)进行系统测试,该数据集是一个 5 分系统的打分记录,包含 6 040 名用户对 3 982 部电影的 1 000 209 条打分记录.具体评测方法如下.

(1)精准度评测方法.本文采用推荐系统中通用的精准度测量方法^[6],首先将数据集分为训练数据集 M 和测试数据集 T , T 是从原有 Movielens 数据集中抽取每个用户的最后一次 5 分记录形成的, T 中只包含 5 分的评分记录,即测试数据集中只包含用户“喜欢”的电影,剩余数据即为 M .对于测试数据集 T 的每条评分记录 (u, i, t, r) ,从电影集合 I 中随机选取 X 个用户 u 尚未观看过的电影,与电影 i 一起进行 top- N 排名,假定电影 i 的最终排名为 e ,如果 $e < N$,则称之为一次命中,即一次成功的推荐.统计所有测试数据集中的命中数目 h ,最终推荐精准度计算公式为

$$p = \frac{h}{N |T|} \tag{6}$$

(2)总体多样性评测方法.总体多样性的评测指标为覆盖率^[2]和系统推荐物品数目^[5],两种评测指标的关系为

$$c = \frac{1}{|I|}d \tag{7}$$

式中: c 为覆盖率; $|I|$ 为系统物品总数; d 为系统推荐物品数目.考虑到两评测指标为线性关系,在结果表达上具有一致性,本文仅选取系统推荐物品数目

作为推荐系统总体多样性的评估指标,计算公式为

$$d = \left| \bigcup_{u \in U} L_N(u) \right| \tag{8}$$

式中: $L_N(u)$ 表示推荐系统为用户 u 推荐的 top- N 物品集合.式(8)表示用户集合 U 中的所有用户接收到推荐结果的不同物品的数目.

d 也可作为一个推荐系统个性化程度的指标,因为一个低的 d 值意味着推荐系统趋于推荐相同的物品,而高的 d 值才说明用户接受的物品各不相同,从而体现了系统推荐结果的个性化.

下面利用 Movielens 的 1 M 数据集,按照上述评测方法对式(1)的标准排名模型 SRM(Standard Rank Model)、文献[5]中的传统排名模型 TRM(Tradition Rank Model)以及本文提出的基于用户评分偏置的排名模型 UBRM(User Bias Rank Model)进行 top-5 推荐测试,测试结果如图 2 所示.

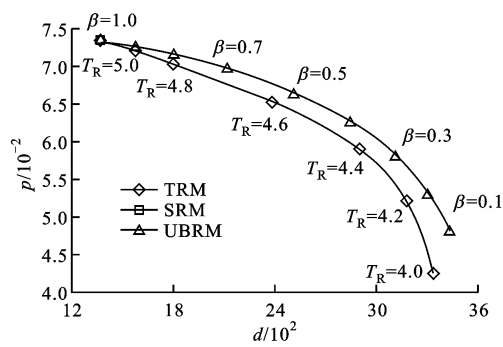


图 2 top-5 排名模型对比测试

两种改进模型中 k_x 都采用物品流行度逆序排名的计算方式,评分预测算法均采用改进的矩阵分

解模型(RSVD)^[15],特征数取 $k=20$,分别记录3种模型的推荐精准度及其对应的多样性结果.当 $T_R=5$ 或 $\beta=1$ 时,两种改进模型都退化为式(1)所示的标准排名模型SRM.从图2中可以明显看到,本文提出的排名模型UBRM相对于文献[5]中的排名模型TRM,在大幅度提高推荐多样性的同时,可以更有效地保证系统的推荐精准度.

进一步地,为了充分说明该模型可以有效控制精准度的下降程度与多样性提升程度的关系,我们分别按物品流行度、物品预测评分逆序、物品平均分、物品的绝对受欢迎程度、物品的评分方差这几种 r_x 的实现方式^[5]对本文的排名模型进行了top-5推荐测试,表1给出了不同精准度下降程度下,各排名

算法总体多样性的相对提升程度.总体多样性的相对提升程度 f_x 和精准度的绝对下降程度 q_x 的计算方法为

$$f_x = \frac{d_x - d_{Std}}{d_{Std}} \times 100\% \tag{9}$$

$$q_x = p_x - p_{std} \tag{10}$$

式中: d_x 、 p_x 为各种排名算法的总体多样性和精准度; d_{Std} 、 p_{Std} 为标准排名算法的总体多样性和精准度,并且有 $d_{Std}=1\,373$, $p_{Std}=0.073\,5$.

在实际应用场景里,可以根据系统需求选择合适的 k_x 计算方式,同时可以通过调节 β 因子,实现精准度和总体多样性的精细控制.

表1 各排名算法总体多样性增量对比

q_x	物品流行度		预测评分逆序		物品平均分		物品绝对受欢迎程度		物品评分方差	
	d	$f/\%$	d	$f/\%$	d	$f/\%$	d	$f/\%$	d	$f/\%$
-0.025 0	3 431	149.9	3 442	150.7	3 406	148.1	3 356	144.4	3 174	131.2
-0.010 0	2 801	104.0	2 951	114.9	2 712	97.5	2 641	92.4	2 674	94.8
-0.005 0	2 287	66.6	2 084	51.8	2 284	66.4	2 036	48.3	2 163	57.5
-0.002 5	1 934	40.9	1 903	38.6	1 798	31.0	1 742	26.9	1 701	23.9
-0.001 0	1 582	15.2	1 440	4.9	1 673	21.8	1 420	3.4	1 524	11.0

3 结 论

随着各种各样的推荐技术运用于推荐系统中,推荐质量得到了大幅度提高.然而,大部分的推荐技术都用于提高推荐的精准度,忽视了推荐结果的总体多样性.因此,本文提出一种加入用户评分偏置的排名模型,在保证推荐精准度的同时,大幅度提高了推荐结果的总体多样性.该排名模型可以方便地嵌入互动电视、电子商务等推荐系统中,无需对原有的推荐功能模块进行改动,排名模型的复杂度趋近于线性,保证了推荐的效率.本文通过在实际的Movielens数据集上对提出的排名模型进行测试,证明了排名模型的实际效用.

参考文献:

[1] MCNEE S M, RIEDL J, KONSTAN J A. Being accurate is not enough: how accuracy metrics have hurt recommender systems[C]//Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems. New York, NY, USA: ACM, 2006:1197-1101.

[2] HERLOCKER J L, KONSTAN J A, TERVEEN L G, et al. Evaluating collaborative filtering recommender systems [J]. ACM Transactions on Information Systems, 2004, 22(1):5-53.

[3] ADOMAVICIUS G, TUZHILIN A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions [J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(6):734-749.

[4] BALABANOVIC M, SHOHAM Y. Fab: content-based, collaborative recommendation [J]. Communications of ACM, 1997, 40(3):66-72.

[5] ADOMAVICIUS G, KWON Y. Improving aggregate recommendation diversity using ranking-based Techniques [C]//Proceedings of the IEEE Transactions on Knowledge and Data Engineering. Piscataway, NJ, USA: IEEE, 2011:1-15.

[6] CREMONESI P, KOREN Y, TURRIN R. Performance of recommender algorithms on TopN recommendation tasks [C]//Proceedings of the 4th ACM Conference on Recommender Systems. New York, NY, USA: ACM, 2010:39-46.

[7] BRADLEY K, SMYTH B. Improving recommenda-

tion diversity [C]//Proceedings of the 12th Irish Conference on Artificial Intelligence and Cognitive Science, Maynooth, Ireland; NUI, 2001;85-94.

[8] SMYTH B, MCCLAVE P. Similarity vs. diversity [C]//Proceedings of the 4th International Conference on Case-Based Reasoning: Case-Based Reasoning Research and Development. Berlin, Germany; Springer-Verlag, 2001;349-361.

[9] ZIEGLER C N, MCNEE S M, KONSTAN J A, et al. Improving recommendation lists through topic diversification [C] // Proceedings of the 14th International Conference on World Wide Web. New York, NY, USA; ACM, 2005;22-32.

[10] ZHANG Mi, HURLEY N. Avoiding monotony: improving the diversity of recommendation lists [C] // Proceedings of the 2008 ACM Conference on Recommender Systems. New York, NY, USA; ACM, 2008;123-130.

[11] MCSHERRY D. Diversity-conscious retrieval [C]// Proceedings of the 6th European Conference on Advances in Case-Based Reasoning. Berlin, Germany; Springer-Verlag, 2002; 219-233.

[12] FLEDER D, HOSANAGAR K. Blockbuster culture’s next rise or fall: the impact of recommender systems on sales diversity [J]. Management Science, 2009, 55 (5):697-712.

[13] ANDERSON C. The long tail-der lange schwanz [M]. New York, NY,USA; Hyperion, 2006;1-15.

[14] GOLDSTEIN D G, GOLDSTEIN D C. Profiting from the long tail [J]. Harvard Business Review, 2006, 84 (6):24-28.

[15] PATEREK A. Improving regularized singular value decomposition for collaborative filtering [C] // Proceedings of the KDD Cup and Workshop. New York, NY, USA; ACM, 2007;39-42.

(编辑 武红江)