

# 基于二分图的自适应推荐算法

刘友林

东华大学信息科学与技术学院 上海 201620

【摘要】本文提出了一种基于赋权二分图的网络结构推荐算法,将归一化后的用户对物品的评分赋值给用户—物品的连边,构建一个用户—物品赋权二分图。提出了一个物品的度调整因子,减少度大的流行物品对其他物品的分类影响。同时,考虑到要减少大量使用流行物品的用户的推荐权重,提高使用非流行物品的用户的推荐权重,提出了用户度调整因子,提高大度用户的推荐能力,因为他们较多选择了非流行物品。测试表明,通过引入用户和物品的自适应度调整因子,提高了推荐算法的精确性。

【关键词】推荐系统 二分图 度 自适应推荐算法

中图分类号:TP3 文献标识码:B 文章编号:1009-4067(2012)01-73-02

## 1. 引言

互联网的迅速发展,将我们带入了海量信息时代。推荐系统作为一种新的获取信息的技术,如今已经出现了很多成熟的算法并成功地应用在商业上<sup>[1]</sup>。当前的推荐系统主要有协同过滤和基于内容过滤的推荐系统,另外最近兴起的基于网络结构的二分图推荐算法<sup>[2]</sup>,是一种新型的协同过滤推荐算法。

## 2. 基于二分图的推荐算法

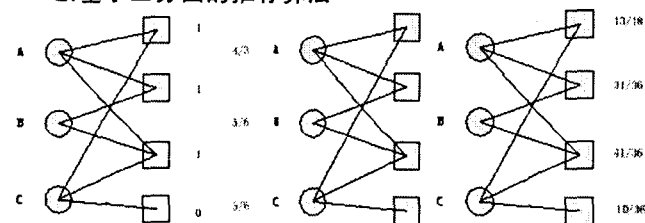


图1 基于二分图的资源分配过程

如上图所示,假设有  $m$  个用户和  $n$  个物品<sup>[3]</sup>,这个系统可以用一个具有  $(m+n)$  节点的二分图表示。对于任意目标用户  $i$ ,推荐算法的目的就是把所有  $i$  没有选择过的物品按照  $i$  喜欢的程度进行排序,并且把排名靠前的那些物品推荐给  $i$ 。假设  $i$  选择过的物品都有向  $i$  推荐某种物品的能力,可以看作位于相关物品上的某种可分的资源。如果用  $w_{ij}$  表示  $j$  与  $i$  的相关程度,可以得到  $w_{ij}$  的表达式:

$$w_{ij} = \frac{1}{k_j} \sum_{l=1}^m \frac{a_{il} a_{jl}}{k_l} \quad (1)$$

其中  $k_j$  和  $k_l$  分别表示物品  $j$  和用户  $l$  的度。用一个  $n$  维的  $0/1$  向量  $f$  表示待测用户的购买记录,代表用户的初始行为  $f^l$ ,记用户  $l$  对应的初始向量为。则有:

$$f^l = W \times f^l \quad (2)$$

向量  $f^l$  是预测的用户对所有物品的评价,  $W$  是由  $w_{ij}$  构成的矩阵。最后将  $f^l$  中用户没有选择过的物品进行排序推荐。

## 3. 基于二分图的自适应推荐算法

在MovieLens数据集中,用户对电影的评分为1到5,用户倾向于对喜欢的电影打较高的分数,反之较低。通过归一化处理后,将用户的具体打分值除以该用户的平均打分值作为相应连边的权值代替原来的无权的0,1模型。这样可以准确地区分用户对该电影的喜好程度,更真实地反映用户行为。

物品的度越大,说明它与其他物品被同一个用户选择的可能性越大。被同一个用户选择,可能是由于此物品比较流行,而并不是因为它与其他物品相关。所以在物品相关性的计算中,应该削弱流行物品对其它物品的推荐能力。文献[4]也通过适当降低流行电影的推荐能力提高了算法的精确性。

用户的度越大,说明这个用户使用过的物品越多,他在系统中的可信度越高。提高大度用户的影响,可以提高系统推荐的准确性和可靠性。O'Mahony等人<sup>[5]</sup>提出了用户概貌的观点,它根据用

户所使用的物品的流行程度来得到。对那些大量使用非流行物品的用户在协同推荐的时候应该加大他们的推荐权重(提高系统鲁棒性和挖掘暗信息的能力);反之则减小。

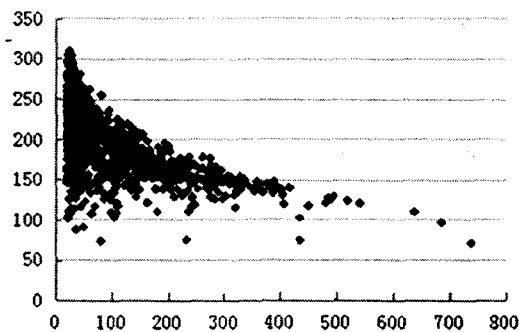


图2 MovieLens数据集中用户概貌曲线分布图

对MovieLens数据集的研究证明了上述观点,即度越大的用户他们选择的物品中包含的非流行物品比例较大;度小的用户倾向于选择流行物品。

算法改进的思想是:在基于赋权二分图的推荐算法中,根据用户的度和电影度的差异,自适应地调整节点之间的信息流。即对(1)式加以修正,得到

$$w_{ij} = \frac{1}{k_j} \sum_{l=1}^m \frac{(a_{il} / \bar{a}_l) \cdot (a_{jl} / \bar{a}_l)}{k_l^{\beta_l}} \quad (3)$$

其中  $\bar{a}_l$  表示用户  $l$  的平均打分值,  $a_{il}$  和  $a_{jl}$  分别表示用户  $l$  对  $i$  和  $j$  的具体打分值。  $\beta_l = \frac{\bar{k}_j}{k_j}$  表示物品  $j$  的度调整因子,等于物品  $j$  的度除以系统中所有物品的平均度,  $\bar{k}_j$  是物品平均度。当  $k_j > \bar{k}_j$ ,说明物品  $j$  较流行,应该降低它的推荐度,反之非流行物品的推荐度则适当提高。  $\beta_l = \frac{\bar{k}_l}{k_l}$  表示用户  $l$  的度调整因子,  $\bar{k}_l$  是用户平均度。当  $\bar{k}_l > k_l$ ,即系统中用户的平均度大于用户  $l$  的度,则此用户的度较小,要降低他在系统中的推荐权重;反之对于度大的用户则需要提高他的推荐影响。

## 4. 实验结果及分析

### 4.1 数据集

我们采用来自明尼苏达州立大学的Group Lens项目组开发的真实数据MovieLens数据集进行研究。Pentium CPU E5300 2.60GHz,内存2G,编程环境为Microsoft Visual C++6.0

### 4.2 推荐结果衡量指标

$\langle r \rangle$  值表示整个系统的平均推荐精度,体现测试集中电影在推荐列表中的相对位置,值越小表明推荐精度越高。初始的基于用户物品二分图网络结构算法(记为算法1)得到的  $\langle r \rangle = 0.106$ ,通过适当降低流行电影的推荐能力(记为算法2)。本文的算法(算法3)得到的  $\langle r \rangle = 0.0915$ ,精度比算法1提高了13.6%,(下转76页)

## 六、对本次#1汽轮机快冷装置应用的分析:

本次快冷装置投运共计48小时,各个金属温度的始终温度和降温率见表3。

表3 投运快速冷却前后主要监视部位金属温度变化

1	部位	始终温度(℃)	平均降温率(℃/h)
2	调节级金属温度	330-107	4.65
3	高压外缸上壁温度	266-105	3.35
4	高压外缸下壁温度	260-85	3.65
5	高压缸上下壁温差	最大 19	< 42
6	中压外缸上壁温度	328-139	3.93
7	中压外缸下壁温度	330-126	4.25
8	中压缸上下壁温差	最大 13	< 42

在冷却过程中,盘车电流无大幅度变化,胀差、汽缸膨胀均匀下降,与自然冷却相比较,投运空气快速冷却后的盘车时间比自然冷却缩小,在16日17点20分停运快冷装置后,盘车继续运行,观察汽缸温度没有回升,19点35分将盘车停运,提前进行汽轮机润滑油滤网的检修。临汾热电#1汽轮机快冷装置的首次投运是成功的,为以后快冷装置的应用积累了宝贵的经验。

对快冷装置投运过程中遇到的一些问题原因分析:

1、最大的降温率是调节级金属温度在开始快冷的前12小时为6.53℃,在之后的24小时、36小时、48小时的降温率分别是5.83℃、4.9℃和4.65℃,冷却是比较均匀的;高压外上缸金属温度的12小时、24小时、36小时、48小时的降温率分别是3.58℃、3.83℃、3.47℃和3.35℃,在24小时的降温率较前12小时快,之后冷却比较均匀,这主要由于高压缸采用压缩空气逆流方式冷却的原因,逆流冷却虽然进口传热温差小,在汽轮机处于高温阶段时受热冲击的风险比顺流冷却小,但因不具备顺流冷却的均匀性优点及进汽区无现成的金属温度测点,而不便于及时监视和调节。中压外上缸金属温度的12小时、24小时、36小时、48小时的降温率分别是3.67℃、4.0℃、3.86℃、3.93℃;中压外下缸金属温度的12小时、24小时、36小时、48小时的降温率分别是4.3℃、4.67℃、4.19℃、4.25℃,从各个时段的降温率看出,中压缸的冷却不均匀,主要原因是中压部分进气管的流通面积比高压部分大。高压内缸、高压隔板套较中压部分壁厚,蓄热量大。

2、内外缸冷却不均匀。#1汽轮机为高、中压缸为合缸结构,没有外部的内外缸夹层冷却管。在汽轮机正常运行中,主蒸汽通过速度级做功后的蒸汽通过高压进汽环的间隙进入内外缸夹层,作为高压部分内外缸夹层的冷却蒸汽;再热蒸汽通过中压进汽环的间隙进入内外缸夹层,作为中压部分内外缸夹层的冷却蒸汽。而在空气快速冷却过程中,进入这部分的压缩空气没有主蒸汽和再热蒸汽那样的压力和流量,也就不能保证有足够的冷却空气通过间隙进入夹层冷却外缸。

3、压缩空气量不足。空气快速冷却系统气源取厂用压缩空气,由于厂用压缩空气用户多而且复杂,不能保证稳定的压缩空气压力,这就限制了汽轮机快冷装置用气量,不能保证达到很好的冷却

效果。

## 七、汽轮机快速冷却的效益分析:

停机过程中300MW机组自然冷却到150℃所需时间在150h以上。300MW机组小修工期18天,大修工期55天,停机后如采用自然冷却,则等待时间占小修工期的1/3以上、占大修工期的1/8。因此减少300MW机组停机后的冷却时间,十分重要。另外,机组故障停机,可采用强迫冷却,使机组以较快的速度冷却降温,尽早开工检修,以减少对限电引起的损失。由此可见,采用快冷装置具有很好的经济效益和社会效益。

## 八、关于寿命损耗的分析:

根据有关资料计算证明,对于汽缸和转子来讲,当其降温率在10℃/小时以内时,通过计算得知,其对汽缸和转子的寿命影响可以忽略不计。我们在设计时是以6℃-8℃/小时以内控制的,这样更不会有大的问题,根据电厂已有的压缩空气量和空气加热器容量而言,一般降温率是不会超过8℃/小时,能控制在6℃/小时左右。即使强迫冷却停机以57℃/小时的温度变化率变化时,它的寿命损耗也只有0.0025%/次。而现场一般快冷的温度变化率在6℃/小时以内,使用一次寿命损耗会更低。由此可见,现场使用快冷对机组寿命的影响可以忽略不计。

## 九、结束语:

在运行人员精心操作下,对#1汽轮机快冷装置进行了成功应用,将缸温在48小时内均匀下降至150℃,为机组尽早检修赢得了时间,更重要的是为年内通过168试运创造了条件,同时这次快冷装置的成功应用为我们积累了宝贵的经验,为将来机组正式运行后快冷装置的再次投用奠定了基础。

## 参考文献

- [1]孙为民,杨巧云主编.《电厂汽轮机》.中国电力出版社
- [2]《汽轮机强迫通风冷却系统说明书》.扬中市华能电力扬设备修造厂
- [3]《临汾热电汽机运行规程》.山西临汾热电有限公司

## 作者简介

裴建东:1972年7月出生,男,山西省长治市人。1993年毕业于太原电力高等专科学校电厂集控运行专业,2001年毕业于太原理工大学电力系统及其自动化专业,工程师。

王宏光:1971年10月出生,男,山西省武乡县人。1992年毕业于太原电力学校发配电专业,2005年毕业于太原电力高等专科学校电力系统及其自动化专业,助理工程师。

董智强,男,1974年出生,1994年毕业于山西大同电力学校热动专业,2001年毕业于山西山西广播电视学校热动专业,技师。

(上接73页)比算法2提高了6%。

## 5. 总结

研究表明,根据节点具体特征,通过自适应地调整节点间信息流分配权值,提高了算法的推荐精度和有效性。

## 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].自然科学进展,2009,1(19):1-15
- [2]Tao Zhou, Ri-Qi Su, Run-Ran Liu, et al. Accurate and diverse recommendations via eliminating redundant correlations[J]. New Journal of Physics, 2009, 11: 123008
- [3]Jian-Guo Liu, Tao Zhou, Bing-Hong Wang, et al. Effects of

User's Tastes on Personalized Recommendation. International Journal of Modern Physics C, 2009, 20: 1925

[4]Tao Zhou, Luo-Luo Jiang, Ri-Qi Su, et al. Effect of initial configuration on network-based Recommendation. Euro physics Letters, 2008, 81: 58004

[5]Michael P. O'Mahony, Neil J. Hurley, and Gurnol C. M. Silvestre. Utility-based neighbourhood formation for efficient and robust collaborative filtering[C]. Ec' 04, May, 2004