

基于时间因子改进个性化推荐模型

胡安明¹, 陈惠娥²

(1.广州理工学院计算机科学与工程学院, 广东 广州 510540;

2.广东金融学院, 广东 广州 510521)

✉anminghu@qq.com; 318802207@qq.com



摘 要: 传统推荐系统算法模型主要集中研究用户偏好与物品的关联性, 根据用户主观意见进行推荐, 未充分考虑用户与物品所处的客观环境, 造成推荐时的实际偏差。本文基于传统推荐算法引入时间因子, 提高模型推荐效果。实现方法主要是通过比较引入与未引入时间因子, 使用UserCF算法和ItemCF算法观察MAE值的大小变化情况。时间因子的引入, 改善了传统推荐系统算法模型主要集中研究用户偏好与物品的关联性等方面的推荐失真问题, 提高了模型推荐的可靠性和实用性。实验结果表明, 引入时间因子能对传统协同过滤算法在MAE指标方面有一定提高, 计算效果优于传统推荐算法。

关键词: 时间因子; 个性化推荐; 协同过滤

中图分类号: TP311.60 **文献标识码:** A

Improved Personalized Recommendation Model based on Time Factor

HU Anming¹, CHEN Huie²

(1.School of Computer Science and Engineering, Guangzhou Institute of Technology, Guangzhou 510540, China;

2.Guangdong University of Finance, Guangzhou 510521, China)

✉anminghu@qq.com; 318802207@qq.com

Abstract: Traditional model of recommendation system algorithm mainly focuses on the relationship between user preferences and items, and makes recommendations according to users' subjective opinions. It fails to take into full consideration the objective environment of the user and the item, resulting in actual deviation in recommendation. This paper proposes to improve model recommendation effect by introducing time factor into traditional recommendation algorithm. The improved model is realized by comparing algorithms with and without time factor, and using UserCF algorithm and ItemCF algorithm to observe the changes of MAE (Mean Absolute Error) values. Introduction of time factor improves the algorithm model of traditional recommendation system, which mainly focuses on the recommendation distortion of user preferences and the relevance of items, so to improve the reliability and practicability of the model recommendation. Experimental results show that introduction of time factor can improve MAE index of traditional collaborative filtering algorithm, and calculation effect is better than that of traditional recommendation algorithm.

Keywords: time factor; personalized recommendation; collaborative filtering

1 引言(Introduction)

近年来, 随着计算机技术的发展和网络的普及, 大量数据信息融入互联网。面对如此巨量的数据信息资源, 如何让用户对其进行精确高效的查询, 有效使用互联网资源; 如何根据用户的个人偏好信息, 结合客观环境, 有效地处理推荐用户所需的信息资源, 仍是目前推荐系统研究的热点。用户—物品间的浏览访问记录和上下文辅助信息数据, 为推荐系统提供了数据挖掘分析的基础^[1], 如何从这些复杂的数据中

挖掘出准确的用户偏好信息, 也是当前研究的热点。

传统的推荐系统模型主要关注用户与物品的行为数据研究, 根据用户的行为数据, 挖掘出用户的特征偏好, 从而进行个人用户的推荐。但这一推荐模式有两种信息未能充分考虑: 其一, 用户和物品所处的客观环境^[2], 例如, 一个食品类型推荐模型, 就不能因为用户有夏秋季节购买冷饮食品的记录, 冬季就继续推荐该类食品。其二, 用户与物品随着时间推移具有普遍的群体性^[3], 例如, 冬季服装主要以各类冬装为

主；外出就餐主要以家庭群组为单位进行。因此，本文提出一种基于时间因子改进的个性化推荐算法模型，从不同的时间点去挖掘用户的喜好，具有较好的实用价值。

2 理论基础(Theoretical basis)

传统的推荐系统模型采用协同过滤算法，该算法的核心思想是：通过用户与物品的历史行为数据，构造协同过滤矩阵^[1]，通过每一行用户对物品的评价，和每一列所有用户对物品的评价，进行协同过滤计算^[2]，从中筛选出用户可能感兴趣的物品。其中，分为UseCF和ItemCF算法。UserCF为基于用户相似度推荐算法，ItemCF为基于物品相似度推荐算法；UserCF适于社交网络推荐应用，ItemCF适于物品推荐；UserCF和ItemCF算法原理基本相同。这里以UserCF算法为例，该算法流程如图1所示。

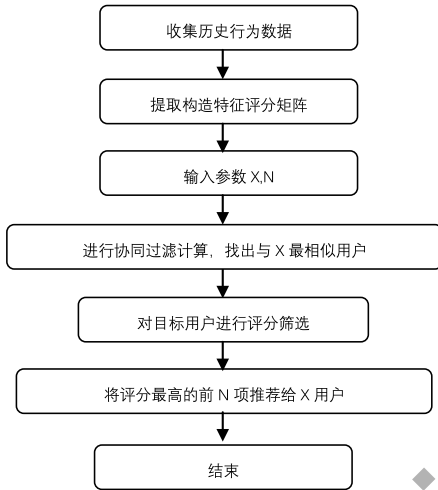


图1 UserCF算法流程

Fig.1 UserCF algorithm flow

协同过滤算法中最为重要的就是用户相似度计算，常用的用户相似度计算算法有余弦相似度、皮尔逊相关系数、改进皮尔逊相关系数算法三类^[4]。

(1) 余弦相似度

余弦相似度计算两个用户向量间的夹角大小，夹角越小，余弦相似度越大，则两个用户越相似。假设数据集中有用户 u 和用户 v 的物品评分，计算用户 u 和用户 v 的相似度，如公式(1)所示；采用集合方式计算，如公式(2)所示，其中 $N(u)$ 为用户 u 点击过的物品评分， $N(v)$ 为用户 v 点击过的物品评分^[5]。

$$\text{sim}(i, j) = \cos(i, j) = \frac{u \cdot v}{\|u\| \cdot \|v\|} \quad (1)$$

$$w_{uv} = \frac{|N(u) \cap N(v)|}{\sqrt{N(u)} \sqrt{N(v)}} \quad (2)$$

(2) 皮尔逊相关系数

皮尔逊相关系数与余弦相似度相比，增加了用户的平均分对数据进行修正，减少了用户评分偏置项的影响^[5]，如公式(3)所示。

$$\text{sim}(i, j) = \frac{\sum_{p \in P} (R_{u,p} - \bar{R}_u)(R_{v,p} - \bar{R}_v)}{\sqrt{\sum_{p \in P} (R_{u,p} - \bar{R}_u)^2} \sqrt{\sum_{p \in P} (R_{v,p} - \bar{R}_v)^2}} \quad (3)$$

其中， $R_{u,p}$ 代表用户 u 对物品 p 的评分， $\bar{R}_v$ 代表用户 v 对所有物品评分的平均分， P 代表所有物品的集合。

(3) 改进皮尔逊相关系数

改进皮尔逊相关系数，将 $\bar{R}_u$ 和 $\bar{R}_v$ 改为物品平均分 $\bar{R}_p$ （物品所有得分的平均分），减少了物品评分偏置对结果的影响^[6]，

如公式(4)所示。

$$\text{sim}(i, j) = \frac{\sum_{p \in P} (R_{u,p} - \bar{R}_p)(R_{v,p} - \bar{R}_p)}{\sqrt{\sum_{p \in P} (R_{u,p} - \bar{R}_p)^2} \sqrt{\sum_{p \in P} (R_{v,p} - \bar{R}_p)^2}} \quad (4)$$

3 基于时间因子改进的推荐模型(Improved recommendation model based on time factor)

传统的推荐算法模型忽视了用户喜好也会随着时间而发生改变。例如，一个食品类型推荐模型，就不能因为用户夏秋季购买过冷饮食品，冬季就继续推荐该类食品。又如，某用户在一时间内对Java语言感兴趣，则这段时间内主要点击的都是与Java相关的图书。但随着时间、工作环境等其他因素的变化，用户又喜欢Python语言了，因而该用户点击的都是Python相关图书。相比较而言，前期的Java语言类图书不能完全代表用户喜好，而推荐系统的目标是推荐用户当前最喜欢的物品，因此，时间因子也是推荐系统模型需要考虑的充分因素。

根据艾宾浩斯遗忘曲线规律^[7]，结合推荐模型中用户和物品行为数据的特征，计算时间衰减函数，如公式(5)所示。

$$\text{time}(t_{u_i} - t_{u_j}) = \frac{1}{1 + \beta |t_{u_i} - t_{u_j}|} \quad (5)$$

其中， t_{u_i} 表示用户 u 对物品 i 发生行为的时间， t_{u_j} 表示用户 u 对物品 j 发生行为的时间； β 为时间衰减因子， β 因子作为阈值，可调节时间的窗口期。

将时间衰减模型引入余弦相似度协同过滤推荐模型^[8]，最近行为与当前推荐物品关系最大，计算如公式(6)所示。

$$w_{uv} = \frac{|N(u) \cap N(v)|}{\sqrt{N(u)} \sqrt{N(v)}} \cdot \text{time}(t_{u_i} - t_{u_j}) \quad (6)$$

在此基础上考虑到不活跃用户大于活跃用户对物品相似度的贡献，降低不活跃用户权重，模型改进如公式(7)所示。

$$w_{uv} = \frac{N(u) \cap N(v) \cdot \frac{1}{\log(1 + N(u))}}{\sqrt{N(u)} \sqrt{N(v)}} \cdot \text{time}(t_{u_i} - t_{u_j}) \quad (7)$$

4 基于时间因子的算法改进流程(Algorithm improvement process based on time factor)

本文引入时间因子算法，相比传统的推荐算法^[9]而言，主要改进在排序阶段，当推荐算法从数据中召回与目标用户可能感兴趣的物品后，在对物品排序过程中引入时间，较好地反映出用户在当前时间段的兴趣特点，体现推荐系统的就近原则^[10]。算法过程如图2所示。

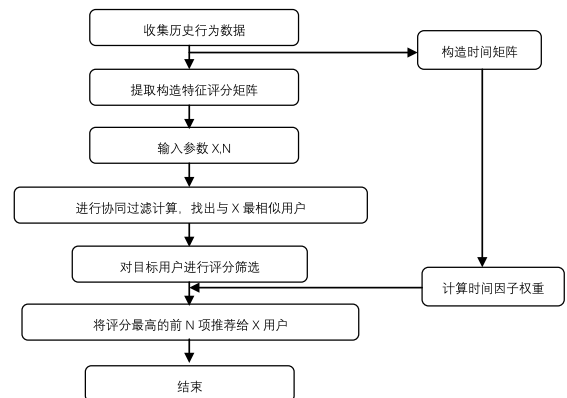


图2 加入时间因子改进的UserCF算法

Fig.2 Improved UserCF algorithm by introducing time factor

5 实验仿真与结果分析(Experimental simulation and result analysis)

本实验软件环境为Windows 10操作系统、Python 3.7, 硬件环境为i5-9400 2.9 GHz、内存16 GB、显卡 GTX1660(6 GB)、硬盘500 GB。实验数据选择MovieLens 1 MB版本数据集, 该数据集中共收录了6,000余名用户对近4,000 部电影的100万条评论数据, 每条评论包含发表评论时间。

本文采用绝对平均误差MAE对算法进行评估, 计算如公式(8)所示, 其中 r_{ui} 为用户 u 对物品 i 的实际评分, $\hat{r}_{ui}$ 为用户 u 对物品 i 的推荐预测。

$$MAE = \frac{1}{n} \sum_1^n r_{ui} - \hat{r}_{ui} \quad (8)$$

本文分别使用UserCF算法和ItemCF算法, 在未引入时间因子情况下和引入时间因子情况下, 相似用户近邻数 n 取值10—70, 分别取七个值。比较引入时间因子和未引入时间因子MAE值的大小变化情况, 其中时间因子超参数 β 取值0.02, 使用结果如表1、图3和图4所示。

表1 MAE值的大小变化情况

Tab.1 Changes in MAE values

用户近邻数 MAE值 计算机模型							
	N=10	N=20	N=30	N=40	N=50	N=60	N=70
UserCF	0.9	0.88	0.84	0.81	0.79	0.78	0.83
引入时间因子UserCF	0.89	0.86	0.82	0.79	0.77	0.79	0.82
ItemCF	0.84	0.8	0.77	0.73	0.71	0.7	0.73
引入时间因子ItemCF	0.83	0.79	0.75	0.71	0.68	0.69	0.71

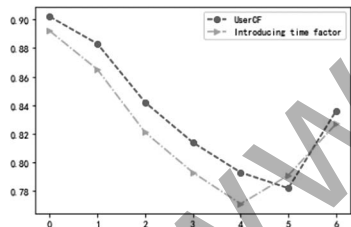


图3 UserCF算法与改进的UserCF算法MAE对比

Fig.3 MAE comparison of UserCF algorithm and improved UserCF algorithm

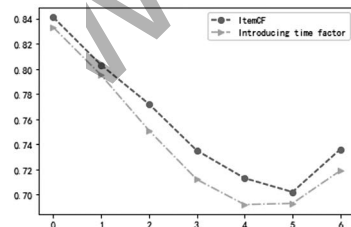


图4 ItemCF算法与改进的ItemCF算法MAE对比

Fig.4 MAE comparison between ItemCF algorithm and improved ItemCF algorithm

时间因子引入过程中, β 为时间因子超参数, 用于控制时间因子time权重值。通过实验分析可知, β 值取值0.02时MAE值最小, 因此可知最近时间用户行为数据具有较大参考价值, 实验结果如图5所示。

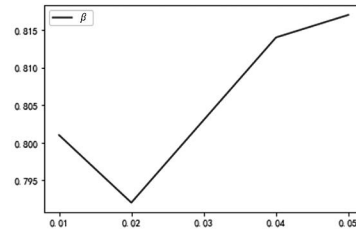


图5 β 超参数取值分析

Fig.5 Analysis of the value of β hyperparameter

实验结果分析: 本文使用MovieLen 1 MB数据集, 在引入时间因子和未引入时间因子情况下, 分别对UserCF算法和ItemCF算法进行测试。在引入时间因子情况下, 实验表明能对传统协同过滤算法在MAE指标方面有一定提高, 计算效果好于传统推荐算法。

6 结论(Conclusion)

本文首先对传统协同过滤算法计算过程进行分析, 在此基础上引入时间因子。通过实验分析, 引入时间因子对传统的协同过滤算法有一定改进效果, 为传统的协同过滤算法的改进提供了一些思路。

参考文献(References)

- [1] 苏庆,章静芳,林正鑫,等.改进模糊划分聚类的协同过滤推荐算法[J].计算机工程与应用,2019,55(5):118-123.
- [2] 王国霞,刘贺平.个性化推荐系统综述[J].计算机工程与应用,2012,48(7):66-76.
- [3] 赵茂先,李小丹.一种带自适应飞行时间因子的粒子群算法[J].山东科技大学学报(自然科学版),2014,33(2):81-85.
- [4] LANDHUIS E.Scientific literature:Information overload[J]. Nature,2016,535(7612):457-458.
- [5] 张玉洁,杜雨露,孟祥武.组推荐系统及其应用研究[J].计算机学报,2016,39(4):745-764.
- [6] 苏庆,陈思兆,吴伟民,等.基于学习情况协同过滤算法的个性化学习推荐模型研究[J].数据分析与知识发现,2020,4(05): 105-117.
- [7] 陶永才,曹朝阳,石磊,等.一种结合时间因子聚类的群组兴趣点推荐模型[J].小型微型计算机系统,2020,41(2):356-360.
- [8] 窦玲源,王新华,孙克.融合标签特征和时间上下文的协同过滤推荐算法[J].小型微型计算机系统,2016,37(1):48-52.
- [9] PESSEMIER T D, DOOMS S, MARTENS L.Comparison of group recommendation algorithms[J]. Multimedia Tools and Applications, 2014, 72(3):2497-2541.
- [10] 陶永才,火昊,石磊,等.基于时间因子的个性化新闻混合推荐研究[J].小型微型计算机系统,2018,39(08):1794-1798.

作者简介:

胡安明(1982—),男,硕士,讲师.研究领域:人工智能,数据挖掘。

陈惠斌(1984—),女,硕士,讲师.研究领域:计算机应用.本文通讯作者。