

一种基于贝叶斯网络的个性化协同过滤推荐方法研究

付永平¹ 邱玉辉²

(安康学院电子与信息工程学院 安康 725000)¹ (西南大学计算机与信息科学学院 重庆 400715)²

摘 要 针对推荐系统不能有效进行个性化推荐问题,在协同过滤过程中引入语义校验,通过对基于用户的协同过滤推荐结果进行语义校验,剔除概率较低的推荐结果,选择概率较高的结果推荐给用户,从而实现个性化语义推荐。在构建贝叶斯语义校验网络时,增加用户“喜好”偏好字段,通过问卷调查及信息反馈,确定用户对物品的喜好偏好值,确保贝叶斯语义校验网络的科学性。实验结果表明,本方法能剔除用户喜好度较低的物品,提高用户的满意度。

关键词 协同过滤,贝叶斯网络,推荐系统,语义

中图分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.9.053

Method of Personalized Collaboration Filter Recommendation Based on Bayesian Network

FU Yong-ping¹ QIU Yu-hui²

(College of Electronic & Information Engineering, Ankang University, Ankang 725000, China)¹

(College of Computer & Information Engineering, Southwest University, Chongqing 400715, China)²

Abstract Lacking of high efficiency personal recommendation in recommendation system, we proposed a new method in collaboration filtering recommendation system by using semantic checking. We check the result of collaboration filtering based on user item by Bayesian semantic to eliminate the item of lower probability, and to select the higher probability item to users. In constructing the Bayesian semantic check network, we add an emotion field named “fancy” by questionnaire survey and information feedback, and we decide user’s emotion for some goods to ensure the scientific of semantic checking network. Experiments show that the method can eliminate the items with low user preferences and improve the satisfaction degree of the users.

Keywords Collaboration filtering, Bayesian network, Recommendation system, Semantic

1 引言

人们迫切需要获取个性化的服务推荐,管理者也亟需高效的推荐系统来改善企业管理、制定科学的决策。目前还缺乏基于用户偏好获取的推荐系统,急切需要根据用户喜好进行推荐的方法和系统。一些大型网站通过消费记录统计出销量大的商品,并结合顾客自己的消费记录向顾客推荐商品。这些商品能被推荐的前提之一是必须要有一定的销量,然而,有一定销量的商品不一定是顾客真正喜欢的商品,因此根据顾客喜好进行研究并进行个性化推荐显得迫切而重要。

当前对推荐系统的理论和研究方法较多,Shu Liao 等, Forsati R 等研究了基于关联规则的推荐系统^[1,2],将当前顾客购买的一系列产品与其他顾客购买的一系列产品作比较,选择顾客购买较多的产品与当前顾客购买的产品集合的交集,最后从中选出一些尚未被顾客所购买但仍然在顾客购物篮中的产品,并将它们作为推荐商品呈现给顾客。Hu Jin-ming, Kim D H, Il Im, Atluri V 等研究了基于内容的过滤技术,根据信息的内容和用户偏好之间的相关性向用户推荐信息^[3,4],它的不足是机器难以自动对内容进行分析,它是一种

基于经验的方法,只能发现与用户已有兴趣相似的信息,而不能为用户发现新的感兴趣的资源。协同过滤推荐技术使得关于个性化推荐系统的研究有了大的进步,P Massa 和 P Avesani 等提出并改进了基于信任的协同过滤系统框架^[5]。为了把情感计算与推荐系统结合起来,以实现根据情感进行语义推荐服务,从而提高推荐的质量, Jinhyung Kim 等, Qi Gao 等利用本体来组织用户和服务信息^[6,7],他们利用本体的描述逻辑描述逻辑是对固有概念的推理,没有结合具体个人研究情感偏好,根据用户的偏好发现用户感兴趣的内容。现实中,人的年龄、性别、民族、宗教、商品本身的属性以及其他顾客购买和使用等情况等都会影响人们的购物行为,这就为个性化情感偏好融入推荐系统提供了市场需求。奉国和等研究了基于 Hadoop 开源分布式计算框架和 Mahout 协同过滤推荐引擎技术构建图书推荐引擎系统,改善了传统单机推荐算法在高维稀疏矩阵上进行运算所导致的系统性能不佳及推荐结果不准确的问题^[8]。孙彦超等从兴趣随时间迁移函数、用户和对象相关函数及用户特性集 3 个方面对传统的协同过滤算法进行优化改进,推荐的准确度得到显著的提高^[9]。武建伟等提出一种基于密度的动态协同过滤图书推荐算法,在提高推

到稿日期:2016-02-14 返修日期:2016-04-29 本文受国家自然科学基金:基于情感语义的全局均衡智能调度理论与方法研究(61152003)资助。

付永平(1965-),男,副教授,主要研究方向为计算机应用,E-mail: akfu@163.com;邱玉辉 男,教授,博士生导师,主要研究方向为人工智能。

荐精确度上,优于传统的协同过滤推荐算法^[10]。Jianhua Liu 提出一种基于简单贝叶斯分类器平滑用户评分等级的个性化信息过滤推荐方法,其可以缓解稀疏性,提高搜索最近邻的准确度^[11]。AS Tewari 等提出了基于内容过滤、协同过滤和关联规则挖掘相结合的图书推荐系统^[12]。

从上述分析可见,当前研究中的主要问题在于个性化语义未能得到体现,依据关联规则来进行推荐只是根据记录数量和品种的关系,发掘的关联规则停留于表面。本体可用来表示语义,但是面对大量用户,构建合适的本体面临极大挑战,本体的构建理论尚未成熟,构建出的本体是否科学有效尚待验证。为了寻找到合适的方法和途径来表示用户的喜好,本研究直接展开问卷调查,并验证调查结果,得到较通用的用户喜好商品数据库,并利用它构建贝叶斯网络,将贝叶斯网络与协同过滤推荐相融合,对协同过滤的结果进行语义校验,力图解决推荐结果的个性化问题。实验结果表明,本方法能剔除用户“喜好”概率较低的推荐商品,把用户“喜好”概率高的商品凸现出来。

2 贝叶斯网络与协同过滤推荐

2.1 贝叶斯网络与贝叶斯公式

贝叶斯网络是一种概率网络,它是基于概率推理的图形化网络,通过一些变量的信息来获取其他的概率信息,以解决不确定性和不完整性问题,在多个领域中获得了广泛应用。而贝叶斯定理是贝叶斯网络的基础,用来描述两个条件概率之间的关系,比如 $P(A|B)$ 和 $P(B|A)$ 。按照乘法法则: $P(A \cap B) = P(A) * P(B|A) = P(B) * P(A|B)$, 可以导出贝叶斯公式: $P(B|A) = P(A|B) * P(B) / P(A)$ 。现实应用中对该公式进行了推广,得到一般贝叶斯公式:

$$P(A_i | B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^n P(B|A_i)P(A_i)}$$

其中, A_1, \dots, A_n 为完备事件组,即, $\bigcup_{i=1}^n A_i = \Omega, A_i \cap A_j = \emptyset, A_j \neq \emptyset, P(A_i) > 0$ 。

2.2 协同过滤(CF)

协同过滤推荐的基本思想是根据用户之前的喜好以及其他兴趣相近的用户的来选择来给用户推荐物品。其基本过程如图 1 所示。

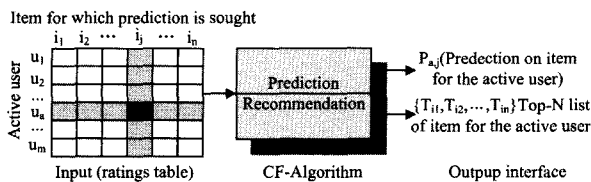


图 1 协同过滤推荐过程

在 CF 中,用 $m \times n$ 的矩阵表示用户对物品的喜好情况,用打分的方式表示用户对物品的喜好程度,分数越高表示越喜欢这个物品,0 表示没有买过该物品。图中一行表示一个用户,一列表示一个物品, U_{ij} 表示用户 i 对物品 j 的打分情况。CF 分为两个过程:预测过程和推荐过程。预测过程是指预测用户对没有购买过的物品的可能打分值;推荐过程是指根据预测阶段的结果推荐用户最可能喜欢的一个或 Top-N 个物品。

基于用户的协同过滤推荐是协同过滤推荐的重要形式,其基本思想是如果用户 A 喜欢物品 a,用户 B 喜欢物品 a, b, c,用户 C 喜欢 a 和 c,那么认为用户 A 与用户 B 和 C 相似,因为他们都喜欢 a,而喜欢 a 的用户同时也喜欢 c,所以把 c 推荐给用户 A。找出相似用户是关键,通常通过计算用户之间的相似度来计算用户相似性。计算相似度的方法较多,如基于余弦的相似度计算、基于关联的皮尔逊相似度计算,本研究采用余弦相似度计算方法。

设 $N(u)$ 为用户 u 喜欢的物品集合, $N(v)$ 为用户 v 喜欢的物品集合,那么 u 和 v 的相似度为:

$$W_{uv} = \frac{|N(u) \cap N(v)|}{\sqrt{|N(u)| * |N(v)|}}$$

3 融合贝叶斯网络的协同过滤推荐模型及算法

3.1 模型

语义校验是个性化推荐的关键,本研究在构建检验数据库时,增加用户对每件物品的情感状态字段,如“喜好”字段,确定用户对该物品的喜好程度。本模型首先获取用户借阅或消费数据库,根据问卷调查及经验,建立贝叶斯校验网络及数据库,然后对用户记录数据库采用协同过滤推荐算法进行推荐,接着用贝叶斯网络对推荐的结果进行校验,剔除不符合要求的推荐结果,得到更可信的推荐结果。协同过滤融合贝叶斯网络的推荐模型如图 2 所示。

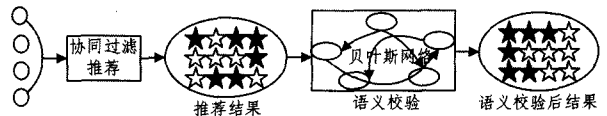


图 2 协同过滤融合贝叶斯网络的推荐模型

3.2 贝叶斯校验协同过滤推荐算法

本模型需要数据库 B_1 , 向用户 R_i 推荐物品 $S_i, S_x \in S_i, N$ 为贝叶斯校验网络, B_2 为校验数据库, U 为用户的基本信息集合, p_x 为推荐的概率。

该算法步骤如下:

Begin:

Step 1 获取 B_1 ;

Step 2 对 B_1 进行协同过滤,得到向用户 R_i 推荐物品的集合 $S_i, S_x \in S_i$;

Step 3 由 B_2 构建贝叶斯网络 N, B_2 由 U 构成;

Step 4 采用余弦相似度公式,在 N 中寻找与 R_i 最相似的用户 $\{R_x\}, x \in [1, \dots, N], x$ 根据需要设定;

Step 5 在 N 中选择 $\{R_x\}$, 将 S_x 作为 N 的输入,得到向 $\{R_x\}$ 推荐 S_x 的概率 $\{p_x\}$, 并计算 p_x 的平均值 \bar{p} ;

Step 6 对 \bar{p} 进行排序,选择值最大的前 m 种物品推荐给用户 R_i 。

End

4 实验

为了验证所提方法的可行性和科学性,用 VC++ 对其编程实现,以大学图书馆向读者推荐图书为例,选取了 3000 条记录,每条记录包括学生的性别、专业、年级、借阅书籍的中图分类号、数量等。贝叶斯网络校验数据库选择了 300 条记录,针对这 300 条记录进行调查验证,以确保该数据库的数据真

实和有效。300 条记录包括专业名称、性别、某类书籍的喜好程度等内容。协同过滤推荐部分结果如表 1 所列。

表 1 协同过滤推荐部分结果

序号	类别	协同过滤结果
1	I5	45.4
2	C9	35
3	I7	31.2
4	I1	27.4
5	I3	26.00001
6	H0	23.10001
7	H1	21.9
8	J2	20.6
9	B2	6.5
10	F0	15.8
11	O1	13.12164
12	B4	0.4
13	P2	0.4

表 1 中,类别指的是图书的分类,序号表示按协同过滤结果排序的排名结果,可见,推荐顺序应该为 I5、C9、I7、I1、I3、H0、H1、J2、B2、F0、O1、B4、P2。

表 2 为进行了贝叶斯网络语义校验后的部分结果,“非常喜欢”、“较喜欢”、“一般喜欢”为读者的个人偏好。在贝叶斯网络数据库中,该值由数据库统计和问卷调查共同得到,表 2 中对应的结果为贝叶斯校验网络校验后的结果。由表 2 可知,贝叶斯网络校验后的推荐顺序与协同过滤的推荐顺序发生了变化,如表 1 中 I5 在协同过滤推荐中排名第一,而校验后,I5 降为第三;而在表 1 排名靠后的 J2、O1 经过贝叶斯语义校验后,推荐顺序上升到第一和第二,对比分析 J2、O1 与读者的专业及年级性别等情况后确定 J2、O1 是读者喜欢的图书。分析发现,单纯的协同过滤推荐的确存在个性化不足的问题,贝叶斯网络语义校验提高了推荐的个性化程度。

表 2 贝叶斯网络语义校验后的部分结果

序号	类别	协同过滤结果	非常喜欢	较喜欢	一般喜欢
8	J2	20.6	0.00032	0.000435	0.000641
11	O1	13.1216	0.00011	0.000435	0.000854
1	I5	45.4	0	0.00065	0.002772
10	F0	15.8	0	0.000435	0
2	C9	35	0	0.000218	0.000214
5	I3	26	0	0	0.000428
7	H1	21.9	0	0	0.000428
4	I1	27.4	0	0	0
6	H0	23.1	0	0	0
9	B2	6.5	0	0	0
12	B4	0.4	0	0	0
13	P2	0.4	0	0	0
3	I7	31.2	0	0	0

结束语 本研究以图书借阅为例,采取了真实的图书借阅数据,构建了校验数据库,在数据库中增加了个性化字段“喜好”,以该数据库为基础,构建了贝叶斯校验网络,利用该网络对协同过滤结果进行了校验,对协同过滤结果进行了重新排名,把读者真正喜欢的图书推荐给用户,增加了读者的满意度。实验表明了该方法的可行性、有效性和科学性。进一步的工作包括继续优化和改进贝叶斯校验网络,比较研究贝叶斯网络校验融入协同过滤的其他途径和方法,从而进一步增加推荐的个性化程度。

参考文献

- [1] Liao S, Zou T, Chang H. An Association Rules and Sequential Rules Based Recommendation System[C] // 4th International Conference on Wireless Communications, Networking and Mobile Computing. IEEE, 2008; 1-4
- [2] Forsati R, Meybodi M, Rahbar R. An efficient algorithm for web recommendation systems[C] // IEEE/ACS International Conference on Computer Systems and Applications. 2009; 579-586
- [3] Hu J. Application and research of collaborative filtering in e-commerce recommendation system[C] // 2010 3rd International Conference on Computer Science and Information Technology (ICCSIT). 2010; 686-689
- [4] Kim D H, Im I, Atlur V. A clickstream-based collaborative filtering recommendation model for e-commerce[C] // IEEE International Conference on E-Commerce. 2005; 84-91
- [5] Massa P, Avesani P. Trust-aware collaborative filtering for recommender systems[C] // Proceedings of International Conference on Cooperative Information Systems. Agia Napa, Cyprus; 2004; 492-508
- [6] Kim J, Jeong D, Baik D K. Ontology-based semantic recommendation system in home network environment[J]. IEEE Transactions on Consumer Electronics, 2009, 55(3); 1178-1184
- [7] Gao Qi, Yan Jun-wei, Liu Min. A Semantic Approach to Recommendation System Based on User Ontology and Spreading Activation Model[C] // International Conference on Network and Parallel Computing. 2008; 488-492
- [8] Feng Gou-he, Huang Jia-xing. Research on Collaborative Filtering Book Recommendation Based on Hadoop and Mahout[J]. Library and Information Work, 2013, 57(18); 116-121 (in Chinese)
- [9] 奉国和, 黄家兴. 基于 Hadoop 与 Mahout 的确协同过滤图书推荐研究[J]. 图书情报工作, 2013, 57(18); 116-121
- [9] Han Feng-xia, Sun Yan-chao. Research on Personalized book Recommendation System Based on Collaborative Filtering Algorithm [J]. Digital Library, 2015(4); 99-102 (in Chinese)
- [9] 韩凤霞, 孙彦超. 基于协同过滤算法的个性化图书推荐系统的研究[J]. 数字图书馆, 2015(4); 99-102
- [10] Wu Jian-wei, Yu Xiao-hong, Chen Wen-qing. Density-based Dynamic Collaborative Filtering Books Recommendation Algorithm [J]. Application Research of Computers, 2010, 27(8); 3013-3015 (in Chinese)
- [10] 武建伟, 俞晓红, 陈文清. 基于密度的动态协同过滤图书推荐算法[J]. 计算机应用研究, 2010, 27(8); 3013-3015
- [11] Liu J. A Personalized Information Filtering Method Based on Simple Bayesian Classifier[M] // Advances in Electronic Commerce, Web Application and Communication. Springer Berlin Heidelberg, 2012
- [12] Tewarii A S, Kumar A, Barman A G, et al. Book Recommendation System Based on Combine Features of Content Based Filtering, Collaborative Filtering and Association Rule Mining[C] // IEEE International Advance Computing Conference. 2014; 500-503