DOI:10.11721/cqnuj20160530

基于贝叶斯网络的个性化关联推荐模型研究*

付永平[1], 胡 勇[2]
(1. 安康学院 电子与信息工程学院, 陕西 安康 725000; 2. 重庆交通大学 信息科学与工程学院, 重庆 400074)

摘要:针对关联规则挖掘不能有效进行个性化推荐问题,研究了关联规则挖掘与贝叶斯网络相融合的个性化关联推荐模型,采用历史记录剪枝与贝叶斯网络校验相结合的办法,对关联规则挖掘算法进行改进。在关联规则挖掘过程中,结合用户历史记录,对关联规则中的频繁项集进行筛选,低于给定阈值项集被剪枝,并把剪枝后的项集输入贝叶斯校验网络进行个性化校验,对校验结果排序后按排名先后进行推荐,实现把读者真正喜欢的图书优先推荐给读者,该推荐模型在一定程度上解决了现有推荐系统中个性化较弱的问题。实验表明,贝叶斯网络可以提高关联推荐的个性化程度。

关键词:贝叶斯网络;关联规则;数据挖掘;推荐系统
中图分类号:TP391 **文献标志码**:A **文章编号**:1672-6693(2016)05-0096-05

人们迫切需要获取个性化的服务推荐,目前还缺乏基于用户偏好获取的推荐系统,一些大型网站通过消费记录,统计出销量大的商品,结合顾客自己消费记录,向顾客推荐商品。这些商品能被推荐,有一个前提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,因此研究根据顾客喜好,并进行个性化推荐显得迫切重要。当前对推荐系统的理论和研究方法较多,Shu 等人研究了基于关联规则的推荐系统[1],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,并将它们作为推荐商品呈现给顾客。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,可以缓解稀疏性,提高搜索最近邻的准确度[2]。张子义等人提出了一种基于关联规则的贝叶斯网络分类算法 BNCAR,该算法利用关联规则挖掘算法提取出初始的候选边,通过贪心算法得到更好的贝叶斯网络结构,BNCAR 能得到较高的分类性能[3]。王刚通过采集顾客观察某物品的眼动参数,借鉴 Find-s 算法思想,提出眼动轨迹语义提取算法。算法通过学习先验知识,让样例正反例距离最大实现确定眼动参数包括注视时间、瞳孔大小、眨眼次数以及回视次数的权重,利用 SEBET 算法,依照距离的远近来判断顾客是否喜欢某商品,实现了从眼动轨迹进行语义提取。该研究没有考虑到个性化关联推荐[4-6]。王礼刚等人提出了一种采用改进型遗传算法的关联规则提取方法,给出了具体的算法,并把这些知识应用到学生的教学管理上去,对原教学制度和计划做出相应的调整。算法在促进学生培养和教育方面有一定的应用价值,该研究未能深入研究用户个性化语义问题[7]。肖海慧等人提出了基于关联规则的 AR-SEM 算法,该算法首先利用关联规则分析变量间的因果关系,并与初始先验知识和领域专家的意见相结合,进一步去除无意义的规则,形成一个知识库,最后将知识库与 SEM 算法相结合来构造贝叶斯网络,该方法虽能在一定程度上提高贝叶斯网络结构学习的精度,但对挖掘出的关联规则未进行个性化检验[8]。在应用方面, Williams 等人建立了一种水坝中鱼腥藻水华的贝叶斯网络模型,模型中,监测数据被存储在一个统一格式的数据库中,通过“学习”因素之间的关系概率,如营养负荷、湖泊水体营养浓度和鱼腥浓度等,能够方便非专业建模者使用,从而显著降低水处理成本和运营支出[9]。Szymon 等人通过海量数据和数据流分析,提出了属性集兴趣度的概念,把属性集特性差异作为数据研究,推导出查找所有属性集的兴趣度超过给定的阈值的精确算法,该算法通过规定相似度和置信概率找到最有趣的属性集[10]。厉海涛等人对贝叶斯网络推理算法研究及近 30 年的发展及功能扩展进行综述,从复杂度、适用性和精度等方面对它们进行比较分析,指出每种算法的关键环节,并对贝叶斯网络在工程技术领域的应用进行了分析回

* **收稿日期**:2016-03-24 **网络出版时间**:2016-07-13 14:00
资助项目:国家自然科学基金 (No. 61152003)
作者简介:付永平,男,副教授,研究方向为计算机应用,E-mail:akfu@163.com
网络出版地址:http://www.cnki.net/kcms/detail/50.1165.N.20160713.1400.004.html

顾，对 BN 的不足和未来的研究趋势做了总结和展望[11-14]。Sahoo 等人针对传统的关联规则挖掘不能反映项目之间的语义量度问题，研究了如何利用效用置信度框架发现关联规则的方法，研究了一种挖掘所有最小前项和最大后项关联规则的密集表示法，并通过高效用闭项集（HUCI）及其发生器实现[15]。在个性化偏好研究方面，Kim 等人利用本体来组织用户和服务信息[16]，根据用户的偏好，发现其感兴趣的内容。

从上述分析可见，当前研究中的主要问题还是个性化语义未能得到充分体现，依据关联规则来进行推荐还停留在根据记录数量和品种的关系，发掘的关联规则实用性不强，不能针对具体用户进行推荐。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。本研究以图书借阅为例，研究个性化推荐的理论和方法，为了体现个性化的针对性，本研究直接采用问卷调查，并进行调查结果验证，得到较通用的能准确真实反映用户喜好图书的数据库，并利用它构建贝叶斯网络，对关联规则挖掘结果进行语义校验，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。

1 基于贝叶斯网络的个性化关联推荐模型

利用关联规则挖掘、历史数据剪枝和贝叶斯网络校验，建立的推荐模型如图 1 所示。该模型包括 4 个功能模块：1）A 模块为关联规则挖掘模块，该模块采用的算法是 Apriori 算法；2）B 模块为历史记录剪枝模块，该模块用历史记录数据对关联规则的进行剪枝，低于给定阈值项集被剪枝；3）C 模块为贝叶斯网络校验模块，用贝叶斯网络对关联规则进行语义校验，对关联项集按概率优先进行排序；4）D 模块为推荐策略制定模块，根据贝叶斯校验网络输出结果，制定出推荐策略。

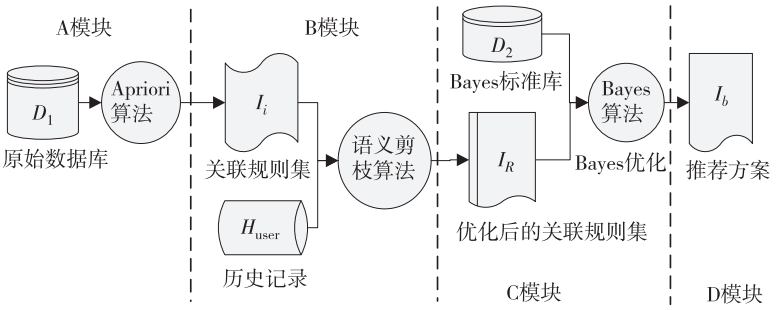

图 1　模型图

Fig. 1　Model diagram

Apriori 算法是一种挖掘关联规则的频繁项集算法，算法分为两个步骤：第一步通过迭代，检索出事务数据库中的所有频繁项集，即支持度不低于用户设定的阈值的项集；第二步利用频繁项集构造出满足用户最小信任度的规则。具体做法就是：首先找出频繁 1-项集，记为 L_1；然后利用 L_1 来产生候选项集 C_2，对 C_2 中的项进行判定挖掘出 L_2，即频繁 2-项集；不断如此循环下去直到无法发现更多的频繁 k-项集为止。

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，以解决应用中一些事实的不定性和不完整性问题，在多个领域中获得广泛应用，而贝叶斯定理是贝叶斯网络的基础。贝叶斯定理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：$P(B|A)=P(A|B)*P(B)/P(A)$，通常对该公式进行推广，得到一般贝叶斯公式：$P(A_i|B)=\dfrac{P(B|A_i)P(A_i)}{\sum_{i=1}^{n}P(B|A_i)P(A_i)}$，其中，$A_1,\cdots,A_n$ 为完备事件组，即 $\bigcup_{i=1}^{n}A_i=\Omega, A_j=\varphi, P(A_i)>0$。

2 个性化关联推荐算法

以图书借阅为例，D_1 为读者借阅记录，算法从 D_1 中进行关联规则挖掘，然后利用用户历史信息对挖掘出的频繁大项集进行剪枝，把剪枝的结果用贝叶斯网络进行校验，得到校验后的频繁大项集，从而制定推荐策略，下面算法中，算法 1 调用了算法 2。

算法 1　贝叶斯个性化关联推荐算法

输入：D_1；

输出：I_b；

Begin

1) 整理读者借阅记录,得到 D_1。

2) 根据 Apriori 算法,设定支持度 Support_degree,生成 k 项集,计算得到要向用户 R_i 推荐内容的大项集 I_i,$I_x \in I_i$, i,$k \in [1,n]$,n 为项目个数。

3) 用户 user 历史数据记录为 H_{user},结合 H_{user},调用 $prune(I_i, H_{\text{user}})$,对 I_i 进行个性化语义剪枝,输出 $I_R = \{I_r | I_r \in I_i\}$。

4) 如果 I_R 满足设定的 k 项集要求,执行 5),否则执行 2),直到满足要求为止。

5) 构建 D_2,D_2 为能反映用户喜欢的借阅记录数据库。

6) 把 I_R 作为 N 的输入,运行 bayes 算法,输出按读者喜好程度排序后的 I_b, $I_b = \{I_y | I_y \in I_i\}$。

7) return (I_b); I_b 即为要向用户 user 推荐的内容。

End

算法 1 中,构建 D_2 的时候,要确保 D_2 的科学性,每条记录包含用户的基本信息为集合 U,用户基本信息有性别、年龄等内容,需根据实际情况,确定用户的基本信息。

本算法对挖掘出来的大项集,利用历史记录进行剪枝,以删除与历史记录吻合度不高的记录。在计算与历史记录吻合度的时候,把大项集与历史记录进行比较,把低于平均值的记录进行剪枝。

算法 2 个性化剪枝算法 $prune(I_i, H_{\text{user}})$

输入:I_i,H_{user}, $I_x \in I_i$;

输出:I_R,$I_R \subset I_i$;

Begin

1) $num = count(I_i)$;num 为大项集项目的个数。

2) For $w = 1$:num

{

3) $T_1 = total(I_w)$; $I_w = \{I_x^w\}$,$I_x^w \in I_x$。

4) $T_2 = count(H_{\text{user}})$; T_2 为 H_{user} 中记录个数。

5) $T_3 = T_1 / T_2$。

6) $T_4 = num / count(H_{\text{user}})$。

7) If($T_3 < T_4$);小于平均值的记录被剪枝。

Delete I_w;从 I_i 中删除 I_x。

}

8) Return (I_R)。

End

算法 2 中 T_4 为本研究选用的阈值。

3 实验

算法用 VC++编程实现,采用图 1 的模型,以学生借阅图书为例,整理了 800 条借阅记录数据库,记录的字段包括学生专业、图书类别、图书名称、学生姓名。对记录库进行关联规则挖掘,选择了某学生,并用其个人历史借阅图书记录进行剪枝。本研究支持度选用 0.05,得到频繁项目集,如表 1 所示。最后采用贝叶斯网络针对项集进行个性化语义校验,得到表 2 所示的结果。对比表 1 和表 2,可看出表 1 挖掘出的关联规则不具有个性化,不同用户的推荐结果相同,这不符合个性化推荐的要求,在表 1 的基础上,采用贝叶斯网络进行个性化语义校验,得到优先推荐的概率排名,从而实现对不同的对象,其推荐结果“不一样”,满足了个性化需求,试验中选择的用户为某名化学专业类学生,得到个性化的结果。关联推荐前 5 名结果为:$I4$,$I5$,$I6$,$I9$;$I1$,$I4$,$I5$,$I9$;$I4$,$I5$,$I7$,$I9$;$I1$,$I4$,$I5$,$I6$;$I4$,$I5$,$I6$,$I7$。要优先推荐的图书类别顺序为{ $I4$,$I5$,$I6$,$I9$, $I7$, $I1$}。通过分析及回访,证明了推荐结果的有效和真实性,即 $I4$,$I5$,$I6$,$I9$, $I7$, $I1$ 是读者真正喜欢的图书。上述实验表明:贝叶斯网络可以提高关联推荐的个性化程度。

表 1　关联规则挖掘结果

Tab. 1　The results of association rule mining

项目集序号	项目集
1	*I*4，*I*5，*I*6，*I*7
2	*I*1，*I*4，*I*5，*I*7
3	*I*1，*I*4，*I*5，*I*6
4	*I*1，*I*5，*I*6，*I*7
5	*I*2，*I*4，*I*5，*I*7
6	*I*4，*I*5，*I*7，*I*9
7	*I*1，*I*2，*I*5，*I*7
8	*I*1，*I*4，*I*6，*I*7
9	*I*2，*I*4，*I*5，*I*6
10	*I*2，*I*5，*I*6，*I*7
11	*I*1，*I*5，*I*7，*I*9
12	*I*1，*I*2，*I*4，*I*5
13	*I*1，*I*2，*I*5，*I*6
14	*I*1，*I*4，*I*5，*I*9
15	*I*4，*I*5，*I*6，*I*9
16	*I*5，*I*6，*I*7，*I*9
17	*I*1，*I*2，*I*4，*I*7

表 2　贝叶斯校验后的结果

Tab. 2　The results of Bayes network checking

优先推荐序号	项目集原序号	项目集	优先度
1	15	*I*4，*I*5，*I*6，*I*9	0.021 452
2	7	*I*1，*I*4，*I*5，*I*9	0.020 113
3	16	*I*4，*I*5，*I*7，*I*9	0.020 111
4	5	*I*1，*I*4，*I*5，*I*6	0.018 705
5	14	*I*4，*I*5，*I*6，*I*7	0.018 703
6	11	*I*2，*I*4，*I*5，*I*6	0.018 618
7	17	*I*5，*I*6，*I*7，*I*9	0.017 826
8	6	*I*1，*I*4，*I*5，*I*7	0.017 364
9	1	*I*1，*I*2，*I*4，*I*5	0.017 280
10	12	*I*2，*I*4，*I*5，*I*7	0.017 278
11	10	*I*1，*I*5，*I*7，*I*9	0.016 487
12	9	*I*1，*I*5，*I*6，*I*7	0.015 078
13	3	*I*1，*I*2，*I*5，*I*6	0.014 994
14	13	*I*2，*I*5，*I*6，*I*7	0.014 992
15	4	*I*1，*I*2，*I*5，*I*7	0.013 654
16	8	*I*1，*I*4，*I*6，*I*7	0.006 874
17	2	*I*1，*I*2，*I*4，*I*7	0.005 449

4 结语

本研究成功把用户历史记录信息与贝叶斯网络校验结合，得到了有效的个性化关联推荐模型，可在一定程度上解决现有推荐系统中个性化较弱的问题。该模型能剔除用户“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，从而把读者真正喜欢的图书优先推荐给读者。进一步的研究包括优化用户历史数据对关联规则挖掘的剪枝方法、改进关联规则的挖掘算法以及对个性化和语义的描述，如引入本体的方法等。

参考文献：

[1] Liao S，Zou T Y，Chang H Y. An association rules and sequential rules based recommendation system[C]//4th international conference on wireless communications，networking and mobile computing. Dalian，China：IEEE Conference Publications，2008(10)：1-4.

[2] Liu J H. A personalized information filtering method based on simple Bayesian classifier[J]. Advances in ECWAC，2012，149：609-614.

[3] 张子义，王德亮. 基于关联规则的贝叶斯网络分类器[J]. 计算机应用，2009，29(6)：134-136.
Zhang Z Y，Wang D L. Bayesian network classifier with association rules[J]. Journal of Computer Applications，2009，29(6)：134-136.

[4] 王刚. 一种基于眼动轨迹的语义提取方法研究[J]. 重庆师范大学学报：自然科学版，2013，30(1)：73-76.
Wang G. Study on method of semantic extraction based on eye tracking[J]. Journal of Chongqing Normal University：Natural Science，2013，30(1)：73-76.

[5] 李梁，张建刚. 基于粗糙集与关联规则的教师科研能力评价[J]. 重庆理工大学学报：自然科学版，2014(1)：69-74.
Li L，Zhang J G. Evaluation of university teacher's research ability based on rough set and association rules[J]. Journal of Chongqing University of Technology：Natural Science，2014(1)：69-74.

[6] 易芝，汪林林，王练. 基于关联规则相关性分析的 Web 个性化推荐研究[J]. 重庆邮电大学学报：自然科学版，2007，19(2)：234-237.
Yi Z，Wang L L，Wang L. Research on Web personalized recommendation based on correlation analysis of association rule[J]. Journal of Chongqing University of Technology：Natural Science，2007，19(2)：234-237.

[7] 王礼刚，左源瑞，李盛瑜. 一种基于改进型遗传算法的关联规则提取算法及其应用[J]. 重庆师范大学学报：自然科学版，2006，23(2)：42-45.
Wang L G，Zou Y R，Li S Y. An algorithm for mining association rules based on improved genetic algorithm and its

application[J]. Journal of Chongqing Normal University: Natural Science,2006,23(2):42-45.

[8] 肖海慧,俞奎,王浩.融合关联规则与知识的贝叶斯网络学习算法[J].微电子学与计算机,2008,25(12):70-72,75.
Xiao H H,Yu K,Wang H. The method of earning Bayesian networks combining association rules with knowledge[J]. Microelectronics & Computer,2008,25(12):70-72,75.

[9] Williams B J,Cole B. Mining monitored data for decision-making with a Bayesian network model[J]. Ecological Modelling,2013,249:26-36.

[10] Szymon J,Tobias S,Dan A S. Scalable pattern mining with Bayesian networks as background knowledge[J]. Data Mining and Knowledge Discovery,2009(18):56-100.

[11] 厉海涛,金光,周经伦,等.贝叶斯网络推理算法综述[J].系统工程与电子技术,2008,30(5):935-939.
Li H T,Jin G,Zhou J L,et al. Survey of Bayesian network inference algorithms[J]. Systems Engineering and Electronics,2008,30(5):935-939.

[12] 黄影平.贝叶斯网络发展及其应用综述[J].北京理工大学学报,2013,33(12):1211-1219.
Huang Y P. Survey on Bayesian network development and application[J]. Transactions of Beijing Institute of Technology,2013,33(12):1211-1219.

[13] 王晓园,蒋经农.贝叶斯线性分层模型估计个体农业保费[J].重庆理工大学学报:自然科学版,2015(9):131-136.
Wang X Y,Jiang J N. Bayesian linear hierarchical models to estimate individual yield crop insurance[J]. Journal of Chongqing University of Technology: Natural Science, 2015(9):131-136.

[14] 王宁,李炜,沈奇威.基于贝叶斯理论的工作流任务分配模型的设计[J].重庆邮电大学学报:自然科学版,2011,23(4):483-486.
Wang N,Li W,Shen Q W. Design of task assignment model for workflow based on Bayesian theory[J]. Journal of Chongqing University of Posts Telecommunications: Natural Science Edition,2011,23(4):483-486.

[15] Sahoo J,Das A K,Goswami A. An efficient approach for mining association rules from high utility item sets[J]. Expert Systems with Applications,2015(42):5754-5778.

[16] Kim J,Jeong D,Baik D. Ontology based semantic recommendation system in home network environment[J]. IEEE Transactions on Consumer Electronics,2009,3(55):1178-1184.

Study on the Personal Association Recommendation Based on Bayes Network

FU Yongping[1], HU Yong[2]

(1. College of Electronic & Information Engineering, Ankang University, Ankang Shaanxi 725000;
2. College of Information Science and Engineer, Chongqing Jiaotong University, Chongqing 400074, China)

Abstract: Lacking of effective personalized recommendation method in association rules discovering, we study a personal model based on Bayes network checking. We use history data pruning and Bayes network checking to optimize the Apriori algorithm, some lower threshold items will be deleted. Pruning lower satisfied degree frequent item sets by using history data, and also to check the personal degree by Bayes network, we do recommendation from sorting order by Bayes probability, which may ensure user to get most satisfied books, the model may solve the problem of lower personal semantic in recommendation system. Experiment show that Bayes network may improve the association recommendation satisfied degree.

Key words: Bayes network; association rule; data mining; recommendation system

(责任编辑 游中胜)