

# 基于行为和评分相似性的关联规则群推荐算法

张佳乐<sup>1</sup> 梁吉业<sup>1,2</sup> 庞继芳<sup>1</sup> 王宝丽<sup>1</sup>

(山西大学计算机与信息技术学院 计算智能与中文信息处理教育部重点实验室 太原 030006)<sup>1</sup>

(太原师范学院计算机系 太原 030006)<sup>2</sup>

**摘要** 使用关联规则推荐工具会遇到最优推荐规则选取难、规则信息不能充分利用等问题。利用较易获取的应用领域知识可有效解决这类问题。针对仅有商品名称和评分信息的推荐情形,提出一种基于行为和评分相似性的关联规则群推荐算法,该算法将规则及相应的评分信息视为推荐专家,将推荐结论相同的专家合并为一个专家组,利用客户行为和评分的双重相似性计算专家权重,并利用群决策的思想集结专家组的推荐意见,从而给出最佳推荐方案。最后通过实例和实验说明了该算法的可行性和有效性。

**关键词** 关联规则,群推荐,行为相似性,评分相似性

**中图法分类号** TP18 **文献标识码** A

## Behavior and Score Similarity Based Algorithm for Association Rule Group Recommendation

ZHANG Jia-le<sup>1</sup> LIANG Ji-ye<sup>1,2</sup> PANG Ji-fang<sup>1</sup> WANG Bao-li<sup>1</sup>

(Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education,

School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China)<sup>1</sup>

(Computer Science and Technology Department, Taiyuan Normal University, Taiyuan 030006, China)<sup>2</sup>

**Abstract** Applying the association rule recommendation tool often meets the hardness of selection for optimal rule, inadequate utilization of rule information. Using easily obtained background knowledge can solve these problems. Aiming at the situation that only contains commodity's name and score information, this paper proposed a behavior and score similarity based association rule group recommendation algorithm, in which the rule with its scores is regarded as an expert, and the experts with same conclusion are grouped together and the expert weights are calculated based on both behavior similarity and score similarity. A better recommendation suggestion is reached by aggregating the recommendation opinions of the experts. The experimental example shows that the algorithm is feasible and effective.

**Keywords** Association rule, Group recommendation, Behavior similarity, Score similarity

## 1 引言

随着电子商务规模的不断扩大、商品个数和种类快速增长,顾客需要花费大量的时间才能找到自己想买的商品。为了解决这个问题,个性化推荐系统应运而生,相关研究逐年递增。然而,由于客户隐私保护问题越来越受到关注,且存在客户填写虚假个人信息等情况,许多需要用到大量客户信息的“奢侈”推荐方法只能被束之高阁。正因为如此,在众多数据挖掘工具中关联分析方法以其简单易用、客户模型直观、可离线完成预计算、不需要分析客户信息及资源内容信息等特点一直被广泛应用于商业推荐系统环境中,尤其是针对电影、音乐、图书等非结构化数据的推荐<sup>[1-3]</sup>。然而,目前应用关联分析方法得到推荐规则集后经常会遇到如下问题:

(1)推荐规则选取难问题。与目标客户匹配的规则可能

有多个,若规则  $r_1$  推荐电影  $M_1$ ,规则  $r_2$  推荐电影  $M_2$ ,而两个规则的各项度量值如支持度和置信度都相同,该选用哪条推荐规则?

(2)规则信息不能全面利用。Thabtah<sup>[4]</sup>曾证明,规则通过支持度和置信度的要求后,都代表了一定的有用知识,不能轻易删除,应该再结合相关领域知识去利用这些知识。例如规则集中有些规则前件会有包含或相交关系,应该如何合理充分利用这些规则信息?

(3)推荐结果满意问题。推荐规则不仅准确率要高而且满意度也要高,所以选择推荐规则不能仅考察支持度、置信度等度量指标,还应充分利用客户反馈的评价信息来保证推荐的满意度。但是这时有可能出现如下的冲突情形:规则  $r_1$  虽然支持度或置信度高于  $r_2$ ,但是规则  $r_2$  的推荐项的评分值高于  $r_1$ ,该选用哪条规则?

到稿日期:2013-05-11 返修日期:2013-09-20 本文受国家 973 计划前期研究专项(2011CB311805),山西省科技攻关计划项目(20110321027-01),山西省科技基础条件平台建设项目(2012091002-0101)资助。

张佳乐(1987—),男,硕士生,主要研究方向为数据挖掘、智能决策,E-mail:15364811828@163.com;梁吉业(1962—),男,教授,博士生导师,主要研究方向为数据挖掘、人工智能、智能决策;庞继芳(1980—),女,博士生,讲师,主要研究方向为数据挖掘、决策理论;王宝丽(1982—),女,博士生,讲师,主要研究方向为智能决策。

针对推荐规则选取问题,较为常用的方法是引入新的度量指标选择最优规则,如规则兴趣度<sup>[5]</sup>、可信度<sup>[6]</sup>、约束规则<sup>[7]</sup>、简单规则<sup>[8]</sup>、最强规则<sup>[9]</sup>等,但这些方法仅就关联规则本身做研究,没有有效利用推荐领域背景知识,可能丢失重要信息。Liu<sup>[10]</sup>指出,与目标客户匹配的多条推荐规则往往从不同角度反映了客户的兴趣和偏好,合理的推荐方法应该结合领域知识综合考虑这些规则的作用。群体决策思想是综合利用信息的有力工具<sup>[11]</sup>。Liu, Jiang<sup>[10,12]</sup>利用群决策思想在集结关联规则时提出将“前件相同,后件不一致”的“冲突”规则看作专家,利用证据理论去集结,并将这一方法运用于电脑购买推荐领域,从客户偏好和满意度之间的关联性角度给出推荐算法,使得规则信息得以充分利用。但是该算法需要从客户文字评论中提取出需求-满意度数据集,获取难度较大,并且需要涉及一些客户个人信息。现实电子商务环境可能是仅有商品名称信息和客户评分信息的情境,如美国 Netflix 公司在 2006 年设立了影片推荐大奖赛,在公开的约 1 亿条影片评级数据中仅包含了影片名称、评价星级和评级日期,希望能用简单的工具在无额外信息的情况下给出尽可能好的推荐。为了解决在仅有商品名称和客户评分信息情形下规则选取难和信息利用不充分等问题,并保证推荐满意度,本文利用群决策思想提出了一种新的关联规则群推荐算法,算法充分利用推荐背景中的客户评分信息,将推荐规则和相应的评分信息视为推荐专家,并将推荐结论相同的专家合成一个专家组,同时考虑行为和评分相似性,设置专家权重,从而集结各专家组的推荐意见,在对组之间推荐意见进行比较后选择最佳推荐方案。本算法推荐的信息利用率高,推荐的结果清晰明了。

## 2 基于行为和评分相似性的关联规则群推荐算法

### 2.1 挖掘推荐规则

设  $I = \{item_1, item_2, \dots, item_n\}$  表示一个项集,其中,  $item_n$  是一个项,表示一个具体的商品。关联规则是形如  $P \rightarrow B$  这样的蕴含式,满足  $P \subset I, B \subset I$ , 并且  $P \cap B = \emptyset$ , 它表示项集  $P$  如果在某一交易中出现,则  $B$  也在同一交易中以某一概率出现,其中  $P$  称为规则前件,  $B$  称为规则后件。  $D = \{t_1, t_2, \dots, t_i, \dots, t_n\}$  表示事务数据库,  $t_i \subset I$ , 一条事务既可以是某客户一次交易中购买的所有商品集合,也可以是某时间阈值内购买的所有商品的集合。

根据实际需要设定支持度和置信度最小阈值,对数据集进行经典关联规则算法挖掘之后,可以得到一个关联规则集  $R$ , 表示如下:

$$R = \{P_1 \rightarrow B_1, P_2 \rightarrow B_2, \dots, P_i \rightarrow B_i, \dots, P_{|R|} \rightarrow B_{|R|}\}$$

$|R|$  是  $R$  中的规则数量,  $P_i \subset I, B_i \subset I, r_i = P_i \rightarrow B_i, P_i = \{item_{i,1}, \dots, item_{i,k}\}, B_i = \{item_{i,k+1}, \dots, item_{i,n}\}$ 。

客户的历史购买记录或点播记录称为客户行为。

目标客户即为需要给予推荐的客户,目标客户  $C_i$  可以表示为  $(P_i, G_i)$ , 其中  $P_i \subset I$  是客户行为,  $P_i = \{item_{i,1}, \dots, item_{i,k}\}, G_i$  为目标客户对  $P_i$  中每个项的评分组成的向量,  $G_i = \{score_{i,1}, \dots, score_{i,k}\}$ 。若目标客户的行为包含某规则的前件,则称该规则与目标客户相匹配,这些规则组成推荐规则集  $R_c, R_c \subset R$ 。生成推荐规则集的具体算法描述如下:

#### 算法 1 生成推荐规则集

输入:目标客户行为  $P_i$ , 规则集  $R$

输出:  $R_c$

$R_c = \emptyset$

For each  $r_i$  in  $R$

If  $P_i \subset P_i$  then  $R_c = R_c \cup r_i // P_i$  是规则  $r_i$  的前件

End for

### 2.2 计算推荐规则评分值

对于推荐规则集  $R_c$  中的每一条规则,它们不仅通过了支持度和置信度的检验,而且与目标客户相匹配,所以它们均具有推荐能力。如果一个客户的购买记录或点播记录满足某一条规则,则可以称这个客户支持这条规则。每一条规则都有若干支持的客户,这些客户对商品项的评分的算术平均值可以作为规则中项的评分值。结合了评分值的推荐规则构成的集合用  $E_c$  表示如下:

设  $E_c = \{(P_1, GP_1) \rightarrow (B_1, GB_1), \dots, (P_j, GP_j) \rightarrow (B_j, GB_j), \dots, (P_n, GP_n) \rightarrow (B_n, GB_n)\}$  为对目标客户  $C_i$  具有推荐能力的规则集合,其中  $n$  为  $E_c$  中规则的个数,  $(P_j, GP_j) \rightarrow (B_j, GB_j)$  是  $E_c$  中第  $j$  条规则,  $P_j$  是规则前件,  $P_j = \{item_{j,1}, \dots, item_{j,k}\}$ 。  $GP_j$  是  $P_j$  中每个项的评分值组成的向量,  $GP_j = \{score_{j,1}, \dots, score_{j,k}\}$ 。  $B_j$  是推荐结论,  $B_j = \{item_{j,k+1}, \dots, item_{j,n}\}$ 。  $GB_j$  是  $B_j$  中每个项的评分值组成的向量,  $GB_j = \{score_{j,k+1}, \dots, score_{j,n}\}$ 。计算推荐规则评分值的算法描述如下:

#### 算法 2 计算推荐规则评分值

输入:推荐规则集  $R_c$

输出:带有评分值的推荐规则集  $E_c$

$C = \{c_1, c_2, \dots, c_{|C|}\}$  是行为满足规则  $r_j$  的客户集,  $E_c = \emptyset$

For each  $r_j$  in  $R_c$

初始化  $r_j$  中每一项的评分属性为空

For each  $item_{j,k}$  in  $r_j$

$$item_{j,k}, score_{j,k} = \frac{1}{|C|} \left( \sum_{i=1}^{|C|} c_i \cdot item_{j,k}, score \right)$$

//其中  $c_i \cdot item_{j,k}, score$  表示客户  $c_i$  的行为记录中项  $item_{j,k}$  的评分值

End for

$E_c = E_c \cup r_j$

End for

### 2.3 生成推荐专家组并集结各组内专家推荐意见

现实中经常遇到这样的决策情形,若干个专家从不同的角度支持同一个观点或者结论,可以将这些专家归到一个组,然后将组内专家对结论的支持程度集结,集结的结果即可代表这个组对该结论的支持程度。在本文中,将带有评分值的推荐规则视为推荐专家,引入如上群决策思想对专家的推荐意见进行集结。

#### 2.3.1 合并推荐结论相同的专家,生成推荐专家组

将推荐专家集  $E_c$  中的推荐结论相同的专家合并为一个推荐专家组,定义如下:

定义 1(推荐专家组)  $GROUP\_E_i$  为  $E_c$  中的任意一个子集,  $GROUP\_E_i = \{(P_{i,1}, GP_{i,1}) \rightarrow (B_{i,1}, GB_{i,1}), \dots, (P_{i,j}, GP_{i,j}) \rightarrow (B_{i,j}, GB_{i,j}), \dots, (P_{i,m}, GP_{i,m}) \rightarrow (B_{i,m}, GB_{i,m})\}$ , 其中  $(P_{i,j}, GP_{i,j}) \rightarrow (B_{i,j}, GB_{i,j})$  是  $GROUP\_E_i$  中的第  $j$  位专家,  $m$  为  $GROUP\_E_i$  中元素的个数。称  $GROUP\_E_i$  为一个专家组,若满足  $B_{i,1} = B_{i,2} = \dots = B_{i,m}$ , 即推荐结论相同。

#### 2.3.2 集结专家组内专家推荐意见

专家的推荐意见包括推荐结论和相应的评分值。集结组

内专家推荐意见之前,首先需要判断组内各专家的权重。专家权重与专家和目标客户的相似性密切相关,相似性越高,专家推荐的可靠性越高。相似性应该考虑两方面内容:行为相似性和评分相似性。行为即购买或点播行为,评分是客户偏好的重要体现,只有行为和评分与目标客户都相似的专家所给出的推荐意见才更加贴近目标客户的喜好。因此,给定专家权重时应该综合考虑这两方面的相似性。

集结组内专家推荐意见的过程描述如下:

(1)计算组内各专家和目标客户的行为相似性

借用文献[13]中的相似性度量方法来计算组内每个专家和目标客户的行为相似性。

定义2 专家组  $GROUP\_E_i$  中专家  $j$  的前件  $P_{i,j}$  与目标客户  $P_t$  的行为相似性定义为:

$$Sim(P_{i,j}, P_t) = |P_{i,j} \cap P_t| \quad (1)$$

归一化处理:

$$BeSim_j = \frac{Sim(P_{i,j}, P_t)}{\sum_{j=1}^m Sim(P_{i,j}, P_t)} \quad (2)$$

其中,  $m$  为  $GROUP\_E_i$  内专家数。

(2)计算组内各专家和目标客户的评分相似性

定义3 专家组  $GROUP\_E_i$  中第  $j$  位专家的评分向量  $GP_{i,j}$  和目标客户的评分向量  $G_t$  的距离为:

$$Dis(GP_{i,j}, G_t) = \frac{1}{|A|} \sum_{item \in A} |(GP_{i,j}, item) - (G_t, item)|$$

其中,  $A = P_{i,j} \cap P_t$ ,  $(GP_{i,j}, item)$  表示  $GP_{i,j}$  在相应  $item$  分量下的  $score$  值。

定义4 专家组  $GROUP\_E_i$  中第  $j$  位专家的评分向量  $GP_{i,j}$  和目标客户的评分向量  $G_t$  的相似性为:

$$ReSim_j = \frac{SIM_j}{\sum_{j=1}^m SIM_j} \quad (3)$$

其中,  $SIM_j = 1 - \frac{Dis(GP_{i,j}, G_t)}{\sum_{j=1}^m Dis(GP_{i,j}, G_t)}$ 。

(3)计算组内专家的权重

定义5 专家组  $GROUP\_E_i$  中专家  $j$  的权重表示为:

$$w_j = \frac{\alpha \times BeSim_j + \beta \times ReSim_j}{\sum_{j=1}^m (\alpha \times BeSim_j + \beta \times ReSim_j)} \quad (4)$$

其中,  $m$  为专家组  $GROUP\_E_i$  中专家的个数。  $\alpha + \beta = 1$ ,  $\alpha$  与  $\beta$  反映的是行为相似性和评分相似性在度量专家权重时的相对重要性,可依实际需要设定。

(4)集结组内专家推荐意见的评分值

每个规则都有支持度和置信度,它们的本质是规则出现的概率,在推荐系统中它反映的是推荐规则的可信度,即相应专家的可信度,所以在集结时也应该予以考虑。

设  $CONF_i = \{conf_1, \dots, conf_j, \dots, conf_m\}$  为专家组  $GROUP\_E_i$  中相应各专家的置信度,  $SUP_i = \{sup_1, \dots, sup_j, \dots, sup_m\}$  为专家组  $GROUP\_E_i$  中相应各专家的支持度。

定义6 专家组  $GROUP\_E_i$  对推荐结论的总评分为:

$$SCORE_i = \sum_{j=1}^m w_j \times GB_{i,j} \times conf_j \times sup_j \quad (5)$$

## 2.4 对专家组的推荐意见进行排序给出最终推荐方案

专家组的推荐意见中项的个数并不相同,用  $|B_i|$  表示专家推荐结论中项的个数,  $|B_i| = 1$  表示推荐一件产品,  $|B_i| > 1$  表示推荐产品组合。最终需要推荐几件产品,要根据实际情况而定。所以,在对专家组排序时,只需要在推荐项个数相同

的组之间进行,总评分最高的专家组的推荐结论就是最终的推荐结果。

参照本节介绍的算法步骤,首先挖掘推荐规则;然后计算推荐规则评分值;接着合并生成推荐专家组并集结各组内专家推荐意见;最后根据系统需求的推荐项个数选择相应的推荐专家组,在选择出的专家组中选择总评分最高的专家组,其推荐结论就是最终的推荐结果。

## 3 算例分析

算例数据如表1所列,商品项集  $I = \{A, B, C, D, E, F, G, H, I\}$ , 客户购买记录中每一商品项下的数字代表客户对该项的评分,评分等级为1至5级。假设目标客户的购买记录是  $\{D, I\}$ , 相应的评分向量是  $(4, 4)$ , 需要据此给他推荐可能最满意的一件商品。

表1 示例数据集

客户 ID	购买记录	客户 ID	购买记录
T1	I, G, F(5, 3, 3)	T6	I, D, E, G, H, A(3, 5, 1, 3, 4, 4)
T2	I, A, C(5, 5, 5)	T7	I, D, G, A(4, 4, 3, 4)
T3	I, C, D, H, A(5, 4, 3, 1, 5)	T8	A, B, E, G(4, 3, 5, 3)
T4	I, G, H, A(4, 3, 4, 4)	T9	A, F, G(4, 5, 3)
T5	D, I, G, H(5, 5, 5, 5)		

Step1 挖掘推荐规则

假设根据需要设定的支持度和置信度的阈值分别是 0.3 和 0.5, 那么使用 Apriori 经典关联规则算法挖掘出规则后, 利用算法1得出与目标客户相匹配的推荐规则, 如表2所列。

表2 推荐规则集

支持度	置信度	推荐规则
0.33	0.75	D, I → A
0.33	0.75	D, I → G
0.33	0.75	D, I → H
0.44	0.57	I → H
0.56	0.71	I → A
0.56	0.71	I → G
0.33	0.75	D → H
0.33	0.75	D → A
0.33	0.75	D → G

表2中, 前3条规则支持度和置信度都相同, 经典的仅基于支持度和置信度选择最优规则的推荐方法就会遇到规则选取困难问题, 随机选取必然会丢失重要信息。实际上, 客户评分是重要的领域知识, 表2后6条规则也都有推荐参考价值, 这些信息都应该被充分利用。所以接下来算法将借用群决策的思想集结规则集, 从而给出最佳推荐方案。

Step2 计算推荐规则评分值

利用算法2对表2的规则集进行处理, 得到带有评分值的推荐规则集  $E_i$ 。也即推荐专家集, 如表3所列。其中, 每个项后面的实数代表这个项的评分值。

表3 推荐专家集

支持度	置信度	推荐专家
0.33	0.75	(D, 4.00), (I, 4.00) → (A, 4.33)
0.33	0.75	(D, 4.67), (I, 4.00) → (G, 4.33)
0.33	0.75	(D, 4.33), (I, 4.33) → (H, 3.33)
0.44	0.57	(I, 4.25) → (H, 3.50)
0.56	0.71	(I, 4.20) → (A, 4.40)
0.56	0.71	(I, 4.20) → (G, 3.80)
0.33	0.75	(D, 4.33) → (H, 3.33)
0.33	0.75	(D, 4.25) → (A, 4.25)
0.33	0.75	(D, 4.67) → (G, 4.33)

### Step3 合并生成推荐专家组并集结各组内专家推荐意见

首先按照定义 1, 将专家集  $E_c$  中各专家进行分组, 生成若干个专家组, 组内专家推荐的结论一致。接下来利用式(1)和式(2)计算组内每个专家和目标客户的行为相似性, 利用式(3)计算组内每个专家和目标客户的评分行为相似性, 利用式(4)计算组内专家的权重, 设  $\alpha=0.5, \beta=0.5$ 。最后, 利用式(5)集结组内评分值, 集结结果如表 4—表 6 所列。

表 4 专家组 G1 推荐意见集结结果

行为相似性	评分相似性	推荐专家	组的总评分
0.5	0.5	(D, 4.00), (I, 4.00) → (A, 4.33)	1.35
0.25	0.28	(I, 4.20) → (A, 4.40)	
0.25	0.22	(D, 4.25) → (A, 4.25)	

表 5 专家组 G2 推荐意见集结结果

行为相似性	评分相似性	推荐专家	组的总评分
0.5	0.36	(D, 4.67), (I, 4.00) → (G, 4.33)	1.22
0.25	0.42	(I, 4.20) → (G, 3.80)	
0.25	0.22	(D, 4.67) → (G, 4.33)	

表 6 专家组 G3 推荐意见集结结果

行为相似性	评分相似性	推荐专家	组的总评分
0.5	0.36	(D, 4.67), (I, 4.00) → (G, 4.33)	1.22
0.25	0.42	(I, 4.20) → (G, 3.80)	
0.25	0.22	(D, 4.67) → (G, 4.33)	

### Step4 对专家组的推荐意见进行排序

专家组的推荐结论中项的个数不一定相同, 我们不必在所有的专家组之间进行排序, 因为需要推荐的产品的个数通常由现实环境决定, 如果推荐项大于 1, 就是推荐商品组合。所以只需要在推荐项个数相同的组之间进行比较即可。本例中系统设定一次只能推荐一个商品, 那么就只需要对推荐项为 1 的专家组的总评分进行排序。比较表 4—表 6 可知, 表 4 所列专家组 G1 的组总评分最高, 所以选择推荐商品 A。

## 4 实验分析

推荐结果有效性判断的常用方法是判断准确率 (Precision)<sup>[14,15]</sup>, 准确率的公式如下:

$$Precision = \frac{\text{准确推荐的客户数}}{\text{测试集中的客户总数}}$$

但是实际中, 不仅要推荐准确, 还应保证推荐结果的客户满意度, 这样的推荐才是成功的推荐, 所以本文定义了推荐满意度 (Satisfaction) 和平均评分值 (Avgscore), 定义如下:

$$Satisfaction = \frac{\text{准确推荐并且满意的客户数}}{\text{准确推荐的客户总数}}$$

$$Avgscore = \frac{\text{准确推荐并且满意的客户评分值之和}}{\text{准确推荐并且满意的客户总数}}$$

本文的实验数据取自 Grouplens Research 公开的电影点播及评分数据集。实验选取的数据集共有 65909 条数据, 是 943 个用户对前 500 部电影的评分记录, 电影评分等级为 1—5 分, 每个用户至少评论 20 部电影。实验按照 75%:25% 的比例随机生成 5 组训练集和测试集, 并且和经典的基于关联规则的推荐算法进行比较, 经典方法通过依次比较规则前件和客户购买记录的匹配度、支持度、置信度来选择最优规则作为最终推荐规则。其中, 客户评分高于 3 分即表示客户对电影满意。实验的运行环境为 Lenovo PC, Intel Core2 Duo

CPU, 2.93GHz, 2G 内存。算法实现采用 Visual basic 语言。实验比较结果如图 1—图 3 所示。

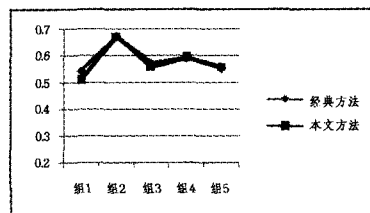


图 1 推荐准确率 Precision 比较结果

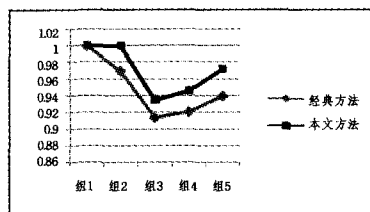


图 2 推荐满意度 Satisfaction 比较结果

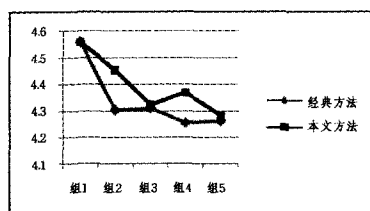


图 3 平均评分值 Avgscore 比较结果

由图 1 可以看到, 经典方法和本文方法的准确率在各组数据集上几乎相同, 但是由图 2 和图 3 所示的满意率和平均评分值的折线走势可知, 本文方法比经典方法整体要高一些。此外, 本文算法的主要计算量在于推荐规则的挖掘和推荐规则评分值的计算, 但是在实际应用中这两步都可以离线完成, 所以完全可以满足网站对响应时间的要求。

**结束语** 公众对隐私问题的敏感导致客户信息越来越难以获取利用, 正因为如此, 一些简单易用、需要信息量少的推荐算法受到了青睐, 关联规则就是其中之一。然而, 使用关联分析方法在推荐后期会遇到最优推荐规则选取难、规则信息不能充分利用等问题, 针对这些问题, 本文提出了一种新的解决方案。本文算法并非直接利用各种度量指标选取最优规则, 删除次优规则, 而是保留所有通过了支持度和置信度要求的有价值规则, 然后利用推荐领域的客户评分信息, 将规则和相应的评分信息视为推荐专家, 利用群决策思想集结各专家推荐意见, 给出推荐方案, 这样不仅避免了规则选取难的问题, 而且充分利用了规则信息以及领域信息, 使推荐过程更加合理有效。最后, 本文选取来自 Grouplens 的数据, 对提出的算法进行分析验证。结果表明, 本文方法的满意率高于经典方法, 验证了本文所提方法的可行性和有效性。实际应用中, 由于数据的复杂性, 单一使用一种推荐工具未必能达到最好的效果, 所以下一步的研究重点是如何将本算法与其它推荐工具结合使用, 使之优势互补, 进一步提高推荐效果 (尤其是准确率), 并且将算法在分布式平台上实现, 从而构建以本算法为核心的高效推荐系统。

## 参考文献

[1] Park D H, Kim H K. A literature review and classification of

- recommender systems research[J]. *Expert Systems with Applications*, 2012, 39(11): 10059-10072
- [2] 哈进兵, 郑锐, 甘利人. 一种基于加权关联规则的协同推荐算法[J]. *情报学报*, 2010, 29(4): 718-722
- [3] 龙舜, 蔡跳, 林佳雄. 一个基于演化关联规则挖掘的个性化推荐模型[J]. *暨南大学学报*, 2012, 33(3): 264-267
- [4] Thabtah F, Cowling P, Peng Y. MMAC: A new multi-class, multi-label associative classification approach[C]//ICDM 2004; Proceedings of the 4th IEEE International Conference on Data Mining, Brighton, UK, 2004: 217-224
- [5] 周欣, 沙朝锋, 朱扬勇. 兴趣度-关联规则的又一个阈值[J]. *计算机研究与发展*, 2000, 37(5): 627-633
- [6] 肖波. 可信关联规则挖掘算法研究[D]. 北京: 北京邮电大学, 2009
- [7] 李广原, 杨炳儒, 周如旗. 一种基于约束的关联规则挖掘算法[J]. *计算机科学*, 2012, 39(1): 244-247
- [8] 杨红菊, 梁吉业. 一种有效的关联规则的挖掘方法[J]. *计算机应用*, 2004, 24(3): 88-89
- [9] 李杰, 徐勇, 王云峰. 面向个性化推荐的强关联规则挖掘[J]. *系统工程理论与实践*, 2009, 29(8): 133-151
- [10] Liu Y Z, Jiang Y C, Liu Y C. CSMC: A combination strategy for multi-class classification based on multiple association rules[J]. *Knowledge-Based Systems*, 2008, 21(8): 786-793
- [11] Pang J F, Liang J Y. Evaluation of the results of multi-attribute group decision-making with linguistic information[J]. *Omega*, 2012, 40(3): 294-301
- [12] Jiang Y C, Shang J, Liu Y Z. Maximizing customer satisfaction through an online recommendation system; A novel associative classification model[J]. *Decision Support System*, 2010, 48(3): 470-479
- [13] Cao F Y, Liang J Y, Li D Y, et al. A dissimilarity measure for the k-Modes clustering algorithm[J]. *Knowledge-Based Systems*, 2012, 26(1): 120-127
- [14] 余力, 刘鲁. 电子商务个性化推荐研究[J]. *计算机集成制造系统*, 2004, 10(10): 1306-1313
- [15] Su J H, Wang B W, Hsiao C Y, et al. Personalized rough-set-based recommendation by integrating multiple contents and collaborative information[J]. *Information Sciences*, 2010, 180(1): 113-131

(上接第 31 页)

- [2] Fang Y, Zio E. Unsupervised spectral clustering for hierarchical modelling and criticality analysis of complex networks[J]. *Reliability Engineering & System Safety*, 2013, 116: 64-74
- [3] Wei L, Keogh E, Xi X. SAXually explicit images: Finding unusual shapes[C]// Sixth International Conference on Data Mining, 2006. ICDM'06. IEEE, 2006: 711-720
- [4] Xi X, Keogh E, Shelton C, et al. Fast time series classification using numerosity reduction[C]// Proceedings of the 23rd international conference on machine learning. ACM, 2006: 1033-1040
- [5] Wang H, Wang W, Yang J, et al. Clustering by pattern similarity in large data sets[C]// Proceedings of the 2002 ACM SIGMOD international conference on management of data. ACM, 2002: 394-405
- [6] Kalpakis K, Gada D, Puttagunta V. Distance measures for effective clustering of ARIMA time-series[C]// Proceedings IEEE International Conference on Data Mining, 2001. ICDM 2001. IEEE, 2001: 273-280
- [7] Aßfalg J, Kriegel H P, Kröger P, et al. Similarity search on time series based on threshold queries[M]// *Advances in Database Technology-EDBT 2006*. Berlin Heidelberg, Springer, 2006: 276-294
- [8] Box G E P. *Time Series Analysis: Forecasting and Control*[M]. Wiley, 2008
- [9] Mörchen F. Time series feature extraction for data mining using DWT and DFT[J]. 2003
- [10] Ji X, Li-Ling J, Sun Z. Mining gene expression data using a novel approach based on hidden Markov models[J]. *FEBS letters*, 2003, 542(1): 125-131
- [11] Jiang J, Zhang Z, Wang H. A new segmentation algorithm to stock time series based on PIP approach[C]// International Conference on Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007. IEEE, 2007: 5609-5612
- [12] Pan S J, Yang Q. A survey on transfer learning[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2010, 22(10): 1345-1359
- [13] Peleg S, Werman M, Rom H. A unified approach to the change of resolution; Space and gray-level[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1989, 11(7): 739-742
- [14] Rachev S T. The Monge-Kantorovich mass transference problem and its stochastic applications[J]. *Theory of Probability & Its Applications*, 1985, 29(4): 647-676
- [15] Hitchcock F L. The distribution of a product from several sources to numerous localities[J]. *J. Math. Phys.*, 1941, 20(2): 224-230
- [16] Rubner Y, Tomasi C, Guibas L J. The earth mover's distance as a metric for image retrieval[J]. *International Journal of Computer Vision*, 2000, 40(2): 99-121
- [17] Pele O, Werman M. A linear time histogram metric for improved sift matching[M]// *Computer Vision-ECCV 2008*. Berlin Heidelberg, Springer, 2008: 495-508
- [18] Pele O, Werman M. Fast and robust earthmover's distances[C]// 2009 IEEE 12th International Conference on Computer vision. IEEE, 2009: 460-467
- [19] Von Luxburg U. A tutorial on spectral clustering[J]. *Statistics and computing*, 2007, 17(4): 395-416
- [20] Malik J, Belongie S, Leung T, et al. Contour and texture analysis for image segmentation[J]. *International Journal of Computer Vision*, 2001, 43(1): 7-27
- [21] Hagen L, Kahng A B. New spectral methods for ratio cut partitioning and clustering[J]. *IEEE transactions on Computer-aided design of integrated circuits and systems*, 1992, 11(9): 1074-1085
- [22] Borgwardt K M, Gretton A, Rasch M J, et al. Integrating structured biological data by kernel maximum mean discrepancy[J]. *Bioinformatics*, 2006, 22(14): e49-e57