

基于 CP-net 偏好的关系数据库的 Top-k 实现

栾艳红, 孙雪姣

(烟台大学计算机与控制工程学院, 山东烟台 264005)

摘要: CP-net 是一种简单而又直观的图形化的偏好表示工具, 能够被用来描述在那些相对严密的、直观的、结构化的条件偏好信息中隐含的定性偏好关系, 尤其适合信息不完整情况下的属性间具有依赖关系的多属性定性偏好决策. Top- k 查询旨在检索出满足用户需求的前 k 个结果, 从而提高检索效率, 为此致力于实现具有 CP-net 偏好的关系数据库的 Top- k 查询. 首先, CP-net 被诱导成多个表来表示和存储; 其次, 将传统的帕累托复合拓展到模型中, 以此保持偏好之间的严格的偏序关系; 最后, 基于“格”框架(Lattice)理论, 实现了基于 CP-net 偏好的关系数据库的 Top- k 查询.

关键词: 条件偏好网; 条件偏好表; 关系数据库; 偏好查询; 偏好复合

中图分类号: TP391 **文献标识码:** A doi: 10.3969/j.issn.0253-2778.2019.02.002

引用格式: 栾艳红, 孙雪姣. 基于 CP-net 偏好的关系数据库的 Top- k 实现[J]. 中国科学技术大学学报, 2019, 49(2):93-99.

LUAN Yanhong, SUN Xuejiao. Top- k query of relational database based on CP-net[J]. Journal of University of Science and Technology of China, 2019, 49(2):93-99.

Top- k query of relational database based on CP-net

LUAN Yanhong, SUN Xuejiao

(School of Computer and Control Engineering, Yantai University, Yantai 264005, China)

Abstract: CP-net is a simple and intuitive graphical preference representation tool that can be used to describe the qualitative preference relationships implied in relatively tight, intuitive, structured conditional preference information. Full-featured qualitative decision-making with multiple dependencies between attributes in complete cases. Top- k queries are designed to retrieve the top k results that meet user requirements, thereby improving search efficiency. Aiming at implement Top- k queries with relational databases that have CP-net preferences. First, CP-net is induced into multiple tables for representation and storage. Then, the traditional Pareto composition is extended into the model so as to maintain a strict partial order relationship between preferences. Finally, based on “Lattice” theory, the Top- k query of relational database based on CP-net preference has been implemented.

Key words: CP-net; CPT; relational database; preference queries; preference composition

收稿日期: 2018-05-24; **修回日期:** 2018-09-18

基金项目: 山东省高等学校科技计划项目(J14LN23, J15LN09); 山东省自然科学基金(ZR2014FL009, ZR2015PF010); 国家自然科学基金(61572419, 61572418)资助.

作者简介: 栾艳红, 女, 1994年生, 硕士生. 研究方向: 数据库. E-mail: 1595247160@qq.com

通讯作者: 孙雪姣, 副教授. E-mail: sunxuejiao6@sina.com

0 引言

用户偏好的需求处理是研究数据库的一个重要方面,因此应该完善在过程查询中关于偏好查询的功能. 之前传统的数据库或者搜索引擎中,查询条件为硬查询,即用户必须给出查询对象的完整信息,系统也只是返回所有满足用户查询条件的查询结果^[1],用户可能面对 2 个问题:

(I) 查询条件的限制导致数据不能完全匹配,查询结果为空集.

(II) 查询条件太弱,导致查询结果太多.

CP-net 适用于描述信息不完全情况下的具有依赖关系的多属性的定性偏好决策,为此现代数据库系统增加了能够通过偏好处理查询的功能,即偏好查询^[2]. 目前为止,在应用搜索引擎和推荐系统中,人们已经开始广泛使用偏好查询. 随着信息技术的广泛应用,各种应用的个性化趋势越来越明显. 例如,新闻浏览系统会根据用户以往的浏览记录为用户推送其他相关新闻,提高新闻浏览量;购物推荐系统会根据用户的购买记录或者浏览记录等为用户定向推荐少量的最可能满足用户要求的商品. 这些个性化应用都需要根据用户的偏好为用户提供最有价值的配置.

在现实生活中,用户的偏好具有以下特点:

(I) 多重属性. 通常,决策的对象可以表示为一组属性. 例如,当用户想要吃饭时,各种属性可能影响着用户的选择:餐厅气氛、价格、主食、饮料、甜点等.

(II) 条件依赖偏好. 某个属性取值的偏好影响用户对其他属性取值的偏好. 例如,汽车颜色的选择取决于车的品牌,主食的选择决定了甜点的种类,这种情况不仅在我们的日常生活中发生,在数据库查询、社会抉择、推荐系统和协同过滤系统,这些多个属性之间存在着依赖关系的定性偏好中也广泛存在.

Top- k 查询的目的就是检索用户给定的相关关键词,并将指定区域分布的最佳 k 个对象提取出来. 在数据库领域中,将偏好集成应用于数据库是一个重要的研究方向,是重要的偏好查询^[3-4]. 其中,Top- k 查询表示数据的排序关系,因此也被称为序敏感查询^[5-6]. CP-net 用于描述在语义中隐藏的、相对紧凑的、直观的、结构化的定性偏好关系. CP-net 的主要优点是一个定性的图形化工具,可以反映属

性间的依赖关系,帮助用户从大量数据中得到自己最关心的配置,提高检索效率.

对于 CP-net,一些详细的语法,语义和应用已经由 Boutilier 等^[7]提出. Wenzel 等^[8]提出了一种结合数据库驱动的推荐方法,利用在线社交网络来寻找具有共同兴趣的人. Sarwat 等^[9]进行了广泛的基于实际的电影推荐和位置感知推荐场景的实验,研究个性化推荐应用的性能. Li 等^[10]和 Hsueh 等^[11]提出预先处理服务,并找到一组可满足用户需求的候选服务要求的服务. Benouaret 等^[12]建议选择一种基于信任的 Top- k 云服务的方法,其中包括信任的提供者和 QoS. Mohammed 等^[13]采用约束 CP-net 对一组数据进行建模,用于表达用户偏好的约束和偏好,这些偏好返回一组列表形式的结果. 这个列表根据用户偏好进行排序,模型不考虑优先级关系,但对查询结果进行量化. 最近,对 CP-net 的子类也进行了一些研究: TCP-net^[14]和 CI-nets^[15],主要添加了反映属性分析 CP-net 的表达力,证明了可满足性和一致性定理. Sun 等^[16]得出结论,CP-net 的一致性等于可满足性,并给出了一致性推理. 最近有用 CP-net 来作为分层表示数据库的偏好建议,通过定义不同优先级的 CP-net. Georgiadis 等^[17]认为,不完备的 CP-net 仅其中的偏好部分即可导致成对元组在某些情况下不能比较. Endres 等^[18]将 CP-net 翻译成偏序表达式的正式偏好语言.

本文的工作不同于现有的方法,从图形角度引入 CP-net 偏好,得出 CP-net 与关系数据库之间的对应关系,CP-net 被诱导和划分成多个数据库表,拓展了帕累托复合. 同时,研究了属性间的偏好关系,并实现了基于 CP-net 的关系数据库的 Top- k 查询. 其中属性关系包括 $f \approx f'$,表示 f 和 f' 是等价的; $f > f'$,表示 f 优于 f' ; $f \perp f'$ 表示 f 和 f' 是冲突关系; $f \parallel f'$ 表示 f 和 f' 的优先级是不能确定的. 本文明确区分在同等优先级和不可比较的元素之间以严格的顺序框架来引出用户偏好表达. 将传统的帕累托复合应用到用户偏好表达中,从而选出最符合用户要求的信息,实现 Top- k 查询.

本文的创新点概括如下:

(I) 计算不依赖于 CP-net 导出图的有效方法,将帕累托复合机制拓展到 CP-net 中.

(II) 采用“格”框架(Lattice)实现具有 CP-net 偏好的关系数据库的 Top- k 查询.

1 基本定义

定义 1.1^[19] 设 $V = \{X_1, X_2, X_3 \dots X_n\}$ 是决策属性的集合, 其中 $\text{Dom}(X_i)$ 代表属性 $X_i \in V$ 的有限定义域 $\{x_1^i, \dots, x_{m_i}^i\}$, 则决策空间 Ω 是各个属性定义域的笛卡尔积, 即 $\Omega = \text{Dom}(X_1) \times \dots \times \text{Dom}(X_n)$ 表示所有属性的可能组合. $o \in \Omega$ 是决策空间的一个配置, 代表决策属性的一种组合. 若两个配置 o 和 o' 仅有一个属性值不同, 其他属性值都相同, 则称 o 和 o' 为可交换的配置. 现实中决策的两个属性之间若有依赖关系, 即决策者对属性 X_i 的偏好取决于 X_j , 则称 X_j 是 X_i 的父亲, 用 $\text{Pa}(X_i)$ 表示.

定义 1.2^[19] $>$ 是决策空间 Ω 上的二元关系, 具有自反性、传递性和反对称性. 当 $>$ 是严格偏序关系时, 则称 $>$ 为 Ω 上的严格偏好关系. $>$ 反映了决策者对两个配置 o 和 o' 的偏好强弱关系, 即 $o' > o$ 表示决策者对 o' 的偏好强于对 o 的偏好.

(I) 若存在两个配置 o 和 $o' \in \Omega$, 有 $o' \not> o$ 且 $o \not> o'$, 则称 o' 和 o 不可比.

(II) 若存在两个配置 o 和 $o' \in \Omega$, 有 $o > o'$ 或 o' 和 o 不可比较, 则含义是 o' 不比 o 强, 表示为 $o >_N o'$. 即 $>_N$ 为 CP-net 所能表达的偏好集合.

定义 1.3^[20] 设 $\text{CPT}(X_i)$ 为属性 X_i 的条件偏好表, 表示属性 X_i 在 $\text{Pa}(X_i)$ 的不同取值下, 决策者对 $\text{Dom}(X_i)$ 为属性 X_i 的集合的一个偏好排序. 由决策者所提供的在 $\text{Pa}(X_i)$ 的一种取值称为属性 X_i 的取值的偏好排序构成了属性 X_i 的条件偏好表 $\text{CPT}(X_i)$.

定义 1.4^[19] CP-net 是一个有向图 $N = \langle V, CE \rangle$, 其中 V 是顶点集, 每一个顶点 X_i 都有一个条件偏好表 $\text{CPT}(X_i)$ 与其关联. CE 为有向边集, 代表所有属性顶点之间的依赖关系, 即一条有向边起点的取值影响终点取值的偏好.

例 1.1 “我的晚餐选择”实例. 晚上决定我的晚餐吃什么, 主要考虑主食、配菜、甜点, 分别用 S, J 和 M 来表示. 我会首先考虑主食的选择, 我对米饭(S_r)的喜爱超过对馒头(S_b)的喜爱. 选择什么样的配菜取决于主食的选择, 如果选择了米饭(S_r), 那么配菜我更偏好于汤类(J_s), 而不是炒菜(J_c); 如果我选择了晚餐主食是馒头, 那么我对炒菜(J_c)比对汤类(J_s)偏好更强. 对配菜的选择又影响了我对甜食的偏好. 对于汤类配菜(J_s), 我更喜欢蛋糕(M_c). 与此相反, 如果选择炒菜配菜(J_c), 我对饼干

甜食(M_b)的偏好要强于对蛋糕甜食(M_c)的偏好.

图 1 表示该例子的 CP-net 图 $N = \langle V, CE \rangle$, 根据题目给出的用户喜好分析得出用户的偏好, $V = \{S, J, M\}$, $\text{Dom}(S) = \{S_r, S_b\}$, $\text{Dom}(J) = \{J_s, J_c\}$, $\text{Dom}(M) = \{M_c, M_b\}$; $CE = \{\langle S, J \rangle, \langle J, M \rangle\}$.

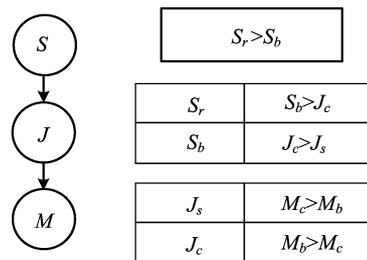


图 1 “我的晚餐选择”CP-net

Fig. 1 CP-net for “my dinner choice”

定义 1.5^[19] 设 $N = \langle V, CE \rangle$ 是一个 CP-net, 则有向图 $G = \langle \Omega, IE \rangle$ 是 N 的导出图, 其中 Ω 是顶点集合, IE 是可交换的配置所构成的有向边集 $G = \langle \Omega, IE \rangle$, 每一条有向边记为 $\langle o_i, o_j \rangle$, 表示由顶点 o_i 指向 o_j , 满足: $o_j > o_i$.

定义 1.6^[19] 给定 CP-net $N = \langle V, CE \rangle$, 其导出图 $G = \langle \Omega, IE \rangle$, 其中 k 为正整数, 则 N 的 Top-k 配置集可定义为: $\text{Top-k}(\Omega) = \{o_i \mid o_i \in \Omega \wedge (1 \leq i \leq k)\} \wedge (\forall o_i \in \text{Top-k}(\Omega)) (\forall o' \in \Omega - \text{Top-k}(\Omega)) (o_i >_N o')$. Top-k 决策指的是在 CP-net 所能表达的所有配置中, 根据属性的占优关系, 确定排在前面的 k 个配置, 所以 Top-k 决策的实质就是要得到 CP-net 的可满足新序列. 比如在浏览新闻网站时, 网站会根据用户的定性偏好信息, 对所有新闻进行排序, 向用户推荐或提供满足用户需求的新闻列表. 满足用户偏好需求的新闻排序可能有多个, 所以有必要求出所有新闻排序, 以满足用户的多元化需求, 更好地实现个性化决策任务.

2 偏好复合

对于人们的多属性偏好, 偏好复合是个性化的核心. 传统的帕累托复合只考虑优先的偏好关系, 而 CP-net 具有更复杂的特性, 所以必须拓展帕累托复合, 如下所示:

$(x, y) > (x', y')$, 当且仅当 $(x <_X x' \wedge y >_Y y') \vee (x >_X x' \wedge y <_Y y')$;

$(x, y) \approx (x', y')$, 当且仅当 $(x \approx_X x') \wedge (y \approx_Y y')$;

$(x, y) \parallel (x', y')$, 当且仅当 $(x \parallel X x') \vee (y \parallel Y y')$;

$(x, y) \perp (x', y')$, 其他情况.

\succ 表示优先的偏好关系, 其他关系还包括 $\approx, \parallel, \perp$. 其中, \approx 是等价的偏好关系, $(x \approx x')$ 表示对于属性 X 来说, x 和 x' 优先级相同; \parallel 表示一种不确定的偏好关系, $(o \parallel o')$ 意味着由于缺少信息, 不能比较这两个结果 (o 和 o'), 也就是说, 我们不能确定究竟是 $o \succ o'$ 还是 $o' \succ o$. 同时, 它也是一个不稳定的关系, 因为 $o \parallel o'$, 可能通过优先关系 \succ 的传递转变成 $o \succ o'$ 或者 $o' \succ o$; \perp 表示一种冲突的关系, $(o \perp o')$ 意味着对于 o 来说, 某些属性的偏好优于 o' , 而有些属性的偏好 o' 优于 o .

定理 2.1 给定两个偏好 $P_x P_y$ 的块序列分别为 $X_0, X_1, X_2, X_3, \dots, X_{l-1}$ 和 $Y_0, Y_1, Y_2, Y_3, \dots, Y_{k-1}$, 那么, 由 $P_x P_y$ 对 $X Y$ 的偏好产生的块序列为 $Z_0, Z_1, Z_2, Z_3, \dots, Z_{l+k-2}$, 包含 $l+k-1$ 块. 每

个块 Z_c 包含了元素块 X_a 和 Y_b , 使得 $a + b = c$.

定理 2.2 给定两个偏好 $P_x P_y$ 的块序列分别为 $X_0, X_1, X_2, X_3, \dots, X_{l-1}$ 和 $Y_0, Y_1, Y_2, Y_3, \dots, Y_{k-1}$, 那么, 由 $P_x P_y$ 对 $X Y$ 的偏好产生的块序列为 $Z_0, Z_1, Z_2, Z_3, \dots, Z_{l+k-1}$ 中, 有 $l * k$ 个组成块. 每个块 Z_c 包含了元素块 X_a 和 Y_b , 使得 $c = a * k + b$. 其中 a 的值范围是从 0 到 $l-1$, b 的值范围是从 0 到 $k-1$, 即 Z_c 的值将从 $X_0 Y_0, X_1 Y_1, X_2 Y_2, X_3 Y_3, \dots, X_{l-1} Y_{k-1}$ 中导出.

例 2.1 给定两个块序列 A_0 和 A_1 及其对应的属性 P_N 和 P_C , 表 1 是用户颜色喜好表, 图 2 中的 N_0 和 N_1, C_0, C_1 分别为 $P_N P_C$ 引起的偏好关系的 4 个块序列. 如图 2(b) 所示, 顶部块 (QB_0) 是 0 元素块的复合索引, 即具有 C_0 的 N_0 , 第二块 (QB_1) 指数块的 sum 为 1, 即具有 C_0 的 N_1, C_1 的 N_0 , 第三块 (QB_2) 为索引和为 2 的块, 即具有 C_1 的 N_1 .

表 1 用户颜色喜好表
Tab. 1 User color preferences

| tid | N | C |
|----------|------|-------|
| T_1 | Tom | blue |
| T_2 | Mary | red |
| T_3 | Mary | blue |
| T_4 | Tom | red |
| T_5 | Jack | white |
| T_6 | Tom | blue |
| T_7 | Jack | black |
| T_8 | Ruth | blue |
| T_9 | Jack | black |
| T_{10} | Tom | black |

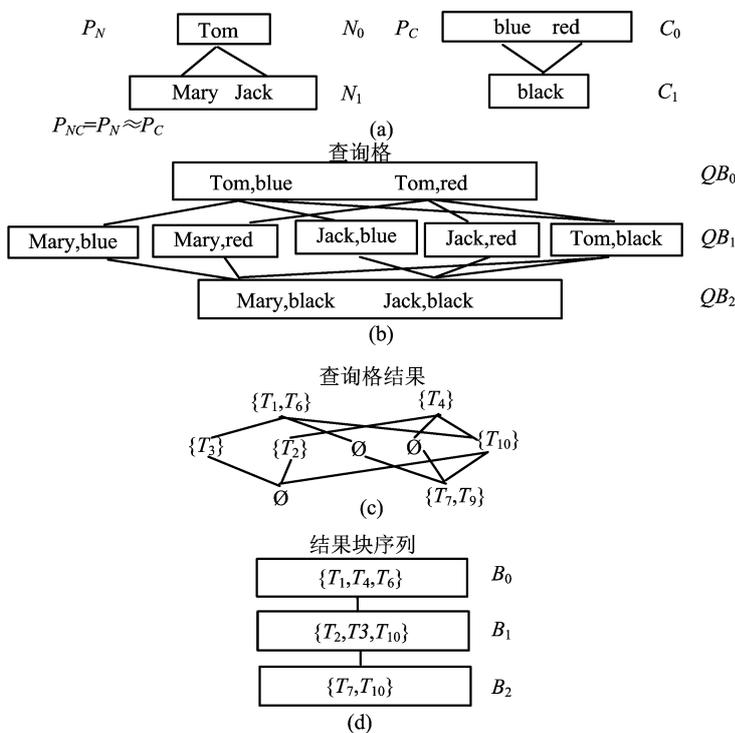


图 2 查询排序框架

Fig. 2 Query sorting framework

本文采用的基本符号 $V(P, A_i)$, 表示的是有效项的优先级集合 PA_i (在图 2 中, $V(P, N) = \{Tom, Mary, Jack\}$). 偏好 P_A 在 R 的属性的非空子集 A 中, $Dom(A)$ 用于表示笛卡尔积 $i(Dom(A_i))$, $V(P, A)$ 表示相应的活动偏好域; 此外, $T(P, A)$ 用于表示该集合 R 的主动元组对每个属性都有活跃

的条件 P_A . 图 2 中, $T(P_{NC}, \{N, C\}) = \{T_1, T_2, T_3, T_4, T_6, T_7, T_9, T_{10}\}$. 再考虑优先表达式:

$$P_{NC} = (P_N \approx P_C), P_N = \{Tom \succ Mary, Tom \succ Jack\}, P_C = \{blue \succ black, red \succ black\}.$$

图 2(b) 显示了诱导偏好 P_{NC} 超过两个活动领域

的笛卡尔积 $V(P, N)$ 和 $V(P, C)$, 它也描绘了 $V(P_{NC}, \{N, C\})$ 上的序列. 计算顶部块 B_0 的偏好查询 PQ_{NC} , 需要执行从查询 $N = \text{Tom} \cup C = \text{blue}$ 和 $N = \text{Tom} \cup C = \text{red}$ 中导出的第一个文档查询块 QB_0 . 由于这两个查询都是非空的结果 ($\{T_1, T_6\}$ 和 $\{T_4\}$ 分别见图 2(c)), 返回关系中唯一的最大元组, 作为顶部 $T(P_N \approx P_C, \{N, C\})$ 的块 B_0 (见图 2(d)), 但并不是格子中的每个查询都是可以保证非空的. 例如, 考虑用户喜好从而获得 T 的下一个块 $(P_N \approx P_C, \{N, C\})$. 如图 2(c) 所示, 在第二格子块 QB_1 的 5 个查询中, 有 2 个空结果, 则其他 3 个非空结果 ($\{T_2, T_3, T_{10}\}$) 属于下一个最大值 B_1 的块 $T(P_{NC}, \{N, C\}) \setminus B_0$. 如果有其他最大值的话, 必须作为后继者的查询结果. 递归地, 我们可以计算 PQ_{NC} 的底部块 B_2 , 如图 2(d) 所示.

2.1 算法实现 Top-k

算法 2.1 构建查询块

输入: 一个偏好表达式 P_A

输出: 查询块序列 QB

```

1  if P is a leaf then
2  QB = PrefBlocks(V(P, Ai))
3  else
4  QB_left = ConstructQueryBlocks(P. left)
5  QB_right = ConstructQueryBlocks(P. right)
6  end if
7  for each w ∈ [0, |QB_left| + |QB_right| - 1] do
8  QB[w] = {QB_left[i] * QB_right[j] | i + j = w}
9  end for
10 return QB

```

算法 2.2 带有 CP-net 偏好的 Top-k 算法实现(SACP)

输入: 一个关系 R, 一个 CP-net 图 $N = \langle V, CE \rangle$

输出: R 的 Top-k 元组

```

1  ConstructQuerySets(Si, Sd)
2  totalize = 0, j = 1
3  QB = ConstructQueryBlocks(PS)
4  i = 0
5  repeat
6  Uqi = GetBlockQueries(QB[i])
7  BSs = Evaluate(Uqi)
8  i += 1
9  until ExitReq or i = |QB|
10 P = BSs
11 Q = Sd
12 N = num(P) // * 块顺序中的块数 * /

```

```

13 repeat
14 M = GetAttribute(Q)
15 repeat
16 P(j) = SubDividedSequ(P(j), CPT(M))
17 j = j + 1
18 until j > N
19 Q = Q - M
20 until Q = ∅
21 Out put(P, k) // * 导出 R 的 Top-k 元组 * /

```

算法 2.2 中 ConstructQuerySets 函数根据集合属性是否有父母将 A 分割成 S_i, S_d . 3~9 行分为两个步骤: 生成查询块序列并输出计算块. 其中, ConstructQueryBlocks(第 3 行)遍历递归地使用偏好表达式树 P_A (来自 P . root), 并计算自下而上的块数将其赋值给 QB. 模型中, $\forall A, B \in S_i$, 因为 S_i 中的属性之间没有依赖关系, 等价偏好关系 $P_A \approx P_B$ 成立. 对于每个 QB 条目, 只有当 \approx 作为表达式 P . left 和 P . right 之间的偏好关系出现时, 才会生成各个块序列的结构; 求取执行其输入集 $\cup q_i$ 的每个查询 q_i 的值(第 7 行). 最后, 输出计算的块并返回其大小. 通过迭代调用 GetBlockQueries(第 6 行)函数来创建关联列表, 连接查询和求值(第 7 行)输出连续的 $T(P, A)$ 块, 可以获得 $T(P_{S_i}, S_i)$ 的连续块, 即 S_i 的块序列 BS. GetAttribute 函数(第 14 行)导出尚未使用的属性组成, 但其所有父属性已经组成 S_d . SubDividedSequ(第 16 行)通过使用属性的 CPT 进一步分割这些块, 逐个分成几个子集. 重复 GetAttribute 和 SubDividedSequ 函数直到组合 S_d 中的所有属性都完成(第 13 到 20 行). 我们可以对 $T(P_A, A)$ 的所有元组进行线性化, 并从中计算出这个序列的 Top-k 元组.

算法的复杂度主要由构造查询块必须执行的查询次数构成. 首先通过遍历查询有关属性的可用索引和 tid (元组标识符)评估, 然后从中获取匹配的元组. 假设为了构造生成的块序列, 共执行 n 个查询 (每个查询为 q), 算法复杂度是 $O(n * (\log |R| + |\text{ans}(q)|))$. 在最好的情况下, 只有一个必需的查询 (即 $n=1$), 返回的元组的数目很小, 特别是当 $|V(P, A)| \ll |T(P, A)|$ 时, 实际复杂度下降到 $O(\log |R|)$. 在最坏的情况下, 为了构造整个块序列, 需要执行所有的格子查询, 实际上几乎返回所有的活动元组, 因此索引遍历的复杂度将上升至 $O(|V(P, A)| * (\log |R|))$, 其中 $|V(P, A)| = \times_i |V(P, A)_i|$, 具体来说, 当 $|V(P, A)| \gg$

$|T(P, A)|$ 时, 算法的实际复杂度成为最坏的情况 $O(|V(P, A)|)$.

3 实例分析

图 3 是购车推荐的 CP-net 模型, 表 2(Numid, P ,

表2 用户买车偏好表
Tab.2 User buy car preference

| Numid | P | B | F | C |
|----------|-------|-------|-------|-------|
| T_1 | P_h | B_b | F_a | C_b |
| T_2 | P_l | B_a | F_b | C_b |
| T_3 | P_t | B_a | F_b | C_w |
| T_4 | P_t | B_b | F_b | C_l |
| T_5 | P_t | B_c | F_a | C_b |
| T_6 | P_l | B_c | F_b | C_l |
| T_7 | P_h | B_a | F_a | C_b |
| T_8 | P_t | B_b | F_a | C_l |
| T_9 | P_l | B_c | F_a | C_l |
| T_{10} | P_h | B_a | F_a | C_w |
| T_{11} | P_h | B_b | F_b | C_w |
| T_{12} | P_l | B_b | F_a | C_b |
| T_{13} | P_h | B_a | F_b | C_w |

B, F, C)描述的是用户买车选择的一部分内容, 为了简单明了, 用元组标识符(Numid)标识每个元组. 以下就是购买车辆用户的购买时的偏好情况:

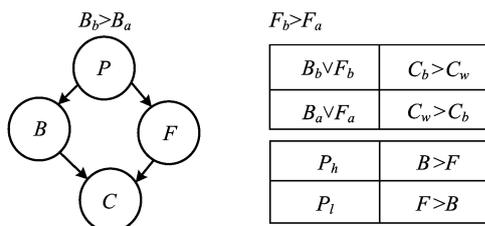


图3 “购车建议” CP-net
Fig.3 CP-net for “Car Advice”

(I) 宝马车的品牌(B_b) 要优于奥迪车的品牌(B_a), 宝马车的整体性能(F_b) 要优于奥迪车的整体性能(F_a).

(II) 当用户选择价位高(P_h) 的车时, 用户考虑车的品牌 B 要优先于考虑车的性能 F .

(III) 当用户同时选择宝马车的品牌(B_b) 和宝马车的性能(F_b) 时, 黑色的车(C_b) 要优于白色的车(C_w).

由表 2 可知, 一共有 4 个不同的属性(即 P, B, F, C). 这是用户进行购车决策时重点关注的 4 个属性. 仅从表 2 中选取依赖于 CP-net 的 Top- k 查询结果, 并推荐给用户.

首先, 要考虑 CP-net 中属性之间的依赖关系, 根据组成偏好关系从简单的块序列生成其他的块序列, 而不需要构建和线性化笛卡尔积, 完成偏好复合的过程. 首先, 按照定理 2.1 “给定两个偏好 $P_x P_y$ 的块序列分别为 $X_0, X_1, X_2, X_3, \dots, X_{l-1}$ 和 $Y_0, Y_1, Y_2, Y_3, \dots, Y_{k-1}$, 那么, 由 $P_x P_y$ 对 $X Y$ 的偏好产生的块序列为 $Z_0, Z_1, Z_2, Z_3, \dots, Z_{l+k-2}$, 包含 $l+k-1$ 块. 每个块 Z_c 包含了元素块 X_a 和 Y_b , 使得 $a + b = c$.” 因此 P, B, F 这 3 块查询块序列的查询块答案如下:

$$\{T_1, T_8, T_{12}\}, \{T_4, T_{11}\}, \{T_7, T_{10}\}, \{T_2, T_3, T_{13}\}.$$

通过使用 CPT(C), 将序列块细分为以下序列:

$$P(1) = \{T_7, T_{10}, T_{13}\} \{T_1, T_2, T_4\} \{T_3, T_8, T_{11}, T_{12}\}.$$

由此, 可以将以下购车选择推荐给用户:

$$\text{Top-3}(R) = \{T_7, T_{10}, T_{13}\}.$$

4 结论

本文主要研究了具有 CP-net 偏好的关系数据库的偏好复合以及 Top- k 的实现过程. 通过分析属性之间的语义关联, 将传统的帕累托复合机制拓展到 CP-net 偏好中, 从而发现隐含的用户偏好, 这样能快速地发现并适应用户兴趣的变化, 保证偏好之间的偏好关系. 在此基础上, 解决了如何使用 CP-net 获取关系表的顶部 k 个元组偏好. 根据属性之间的偏好关系计算并返回符合该用户需求的前 k 个结果. 作为一个崭新的领域, Top- k 查询仍有很多工作需要展开:

(I) Top- k 的焦点问题是探究能够使 Top- k 从语义到返回结果走向统一的更合理的平衡方式.

(II) 将 Top- k 问题扩展到更丰富的数据模型上进行查询处理.

(III) 将 Top- k 应用到 P2P 环境下的数据管理系统中.

参考文献(References)

- [1] GEORGIADIS P, KAPANTAIKIS I, CHRISTOPHIDES V, et al. Efficient rewriting algorithms for preference queries [C]// International Conference on Data Engineering. Cancun, Mexico: IEEE, 2008: 1101-1110.
- [2] SUN Xuejiao, LIU Jinglei, WANG Kai. Operators of preference composition for CP-nets [J]. Expert Systems with Applications, 2017, 86(15): 32-41.
- [3] GIRARD P. Modallogics for belief and preference change[D]. Palo Alto: Stanford University, 2008.
- [4] KIEBLING W. Foundations of preference in database systems[C]// Proceedings of the 28th International Conference on Very Large Data Bases. Hong Kong, China: ACM, 2002:311-322.
- [5] Ilyas I. F. Rank-aware query processing and optimization [D]. West Lafayette: Purdue University, 2004.
- [6] 邓波. 分布式序敏感查询处理关键技术研究[D]. 长沙: 国防科学技术大学, 2006.
- [7] BOUTILIER C, BRAFMAN R I, DOMSHLAK C, et al. CP-nets: A tool for representing and reasoning with conditional ceteris paribus preference statements[J]. Journal of Artificial Intelligence Research, 2011, 21(1): 135-191.
- [8] WENZEL F, KIEBLING W. A preference-driven database approach to reciprocal user recommendations in online social networks [C]// 27th International Conference on Database and Expert Systems Applications. Switzerland: Springer, 2016: 3-10.
- [9] SARWAT M, MORAFFAH R, MOKBEL M F, et al. Database system support for personalized recommendation applications[C]// 33rd International Conference on Data Engineering. San Diego, USA: IEEE, 2017: 1320-1331.
- [10] LI J, YAN, Y LEMIRE D. Scaling up web service composition with the skyline operator [C]// International Conference on Web Services. San Francisco, USA: IEEE, 2016: 147-154.
- [11] HSUEH Y L, LIN C C, CHANG C C. An efficient indexing method for skyline computations with partially ordered domains[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(5): 963-976.
- [12] BENOURET K, BENOURET I, BARHAMGI M, et al. Top- k cloud service plans using trust and QoS [C]// International Conference on Services Computing. Honolulu, USA: IEEE, 2017: 507-510.
- [13] MOHAMMED B, MOUHOUB M, ALANAZI E. Combining constrained CP-nets and quantitative preferences for online shopping [C]// 28th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems. Tokyo, Japan: IEA/AIE, 2015: 702-711.
- [14] BRAFMAN R I, DOMSHLAK C, SHIMONY S E. On graphical modeling of preference and importance [J]. Journal of Artificial Intelligence Research, 2011, 25(1): 389-424.
- [15] BOUVERET S, ENDRISS U. Conditional importance networks: A graphical language for representing ordinal, monotonic preferences over sets of goods [C]// International Joint Conference on Artificial Intelligence. Pasadena, USA: Morgan Kaufmann, 2009: 67-72.
- [16] SUN X J, LIU J L. On the satisfiability and consistency for CP-nets [J]. Journal of Computer Research & Development, 2012, 49(4): 754-762.
- [17] GEORGIADIS P, KAPANTAIKIS J, CHRISTOPHIDES V, et al. Efficient rewriting algorithms for preference queries[C]// International Conference on Data Engineering. Cancun, Mexico: IEEE, 2008: 1101-1110.
- [18] ENDRES M, KIEBLING W. Transformation of TCP-net queries into preference database queries [C]// European Conference on Artificial Intelligence. Hague, Holland: 2006-advances in preference handling.
- [19] 孙雪皎, 刘惊雷. CP-nets 的可满足性序列求解算法研究[J]. 计算机科学, 2015, 42(5): 270-273.
- [20] 惊雷, 廖士中, 张伟. CP-nets 的完备性及一致性研究[J]. 软件学报, 2012, 23(6): 1531-1541.