理时间可能差异较大，导致管理节点对各计算节点返回数据进行综合的时间延误增大，降低了计算节点 CPU 的利用率.为了解决上述问题，在任务划分的基础上进一步细分.令 FCD 处理的最大时间差异为 Δ，则 FCD 任务的最终划分数为

$$N = N_0 \times \Delta \tag{2}$$

假设在一次动态交通诱导处理过程中可控制的空闲 CPU 数量不变，且用户指定的 FCD 处理的约束时间 T_{res} 可满足，给出 FCD 并行计算的动态任务划分与调度算法如下：

算法 2.1　动态任务划分与调度

Step 1：初始化，令一个浮动车数据计算的最大时间为 T_{max}，空闲的计算节点 CPU 数为 C_{idle}，$C_{idle}(i)$ 为第 i 次迭代时空闲的计算节点 CPU 数，FCD 处理的约束时间 T_{res}，FCD 的大小为 M，则根据式 2 计算 FCD 包的划分组数为 N；

Step 2：将电子地图数据传给 C_{idle} 个计算节点 CPU，C_{idle} 份浮动车数据传给对应的 CPU，$M \leftarrow M - C_{idle} \times M/N$，$N \leftarrow N - C_{idle}$，迭代次数 i=1；

Step 3：While (N>0) do

{管理节点监测是否有计算结果从计算节点 CPU 返回，如有则转下一步，否则继续监测；

if $N < C_{idle}(i)$ and $2N \leqslant C_{idle}(i)$

{将剩余的 N 组浮动车数据均分为 2N 组，

$N \leftarrow 2N$}

将 $C_{idle}(i)$ 组浮动车数据分别分配给 $C_{idle}(i)$ 个计算节点 CPU；$N \leftarrow N - C_{idle}(i)$；$i = i+1$}

Step 4：综合各计算节点 CPU 的数据，对路网各路段的速度、密度和流量进行推理，传送至 Web 服务器；

Step 5：进行下一批 FCD 流数据处理.

3　实验结果和讨论

基于本文的方法，我们以国内首台龙芯国产大数据一体机为平台，建立了 FCD 并行处理系统.该大数据一体机由 12 个龙芯 4 核 3B2000CPU 组成的小规模集群机，为 FCD 的并行计算提供了有效的平台.我们使用现场 FCD 数据对本文方法进行验证，并与轮询调度算法和 Min-Min 算法进行比较.该数据为北京 2 万辆出租车的 FCD，各浮动车的采样周期为 15 秒，每 3 分钟向远程中心传送一组 FCD 纪录.算法的评价指标采用加速比和效率.加速比和效率能够衡量一个并行算法性能的优劣.给定一个待求解问题，设 T_p 是使用 P 台处理器并行算法求解的运行时间，T_1 是使用单台处理器最快的串行算法求解运行时间，则使用 P 台处理器并行算法对使用单台处理器串行算法的加速比 $S_p = T_1/T_p$，效率 $E_p = S_p/P$.实验结果如图 2，3 所示.

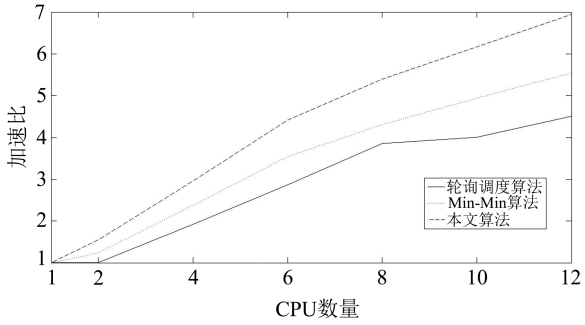


图 2　加速比随 CPU 数量变化曲线

Fig.2　The curves of speedup change with the number of CPU

从图 2 所示结果可以看出，本文算法的加速比明显优于轮询调度算法和 Min-Min 算法，其加速比和计算节点数量之间存在接近线性的关系，其加速比较 Min-Min 算法、轮询调度算法分别提高了约 20％和 35％.

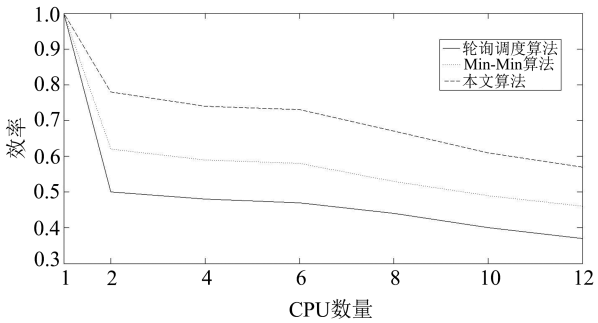


图 3　效率随 CPU 数量变化曲线

Fig.3　The curves of efficiency change with the number of CPU

图 3 显示，当本文算法用于 FCD 大数据并行处理时，其效率优于轮询调度和 Min-Min 算法.当 CPU 数量增加至 12 个时，其效率一直高于 55％.

进一步比较了 Min-Min 算法和本文所提算法用来进行 FCD 处理时 CPU 使用情况.结果如表 1 所示.其中所处理的浮动车数据取 3 分钟时间段的实时 FCD，为便于比较，CPU 数固定为 12 个，任选部分浮动车数据进行处理.

Min-Min 算法的 CPU 平均等待时间为 1.1 秒、平均利用率为 92.8％；本文提出的动态任务调度算法的 CPU 平均等待时间为 0.54 秒、平均利用率为 95.2％，运行结果验证了本文所提方法的有效性.

表 1 CPU 使用状况(时间单位:秒)

Tab.1 The CPUs usage(Unit: s)

CPU 数量	执行时间(sec)(Min-Min 算法)	等待时间(sec)(Min-Min 算法)	执行时间(sec)(本文算法)	等待时间(sec)(本文算法)
01	9.27	1.115	9.507	0.563
02	9.561	0.820	9.233	0.828
03	8.455	1.924	9.923	0.137
04	7.625	2.76	9.507	0.552
05	10.244	0.137	10.037	0.021
06	9.152	1.23	10.057	0.007
07	9.956	0.423	9.507	0.552
08	9.702	0.683	10.063	0.01
09	9.702	0.683	9.233	0.828
10	9.427	0.957	9.233	0.828
11	9.286	1.093	8.958	1.101
12	9.018	1.367	8.958	1.101

4 结论

交通信息获取是交通规划、管理和调控的基础,FCD 大数据处理是获取大范围路网交通实时状态的有效手段.针对由于噪声干扰导致 FCD 采样间隔的不确定性,本文提出了 FCD 大数据任务动态划分和调度算法,并以龙芯国产大数据一体机为平台进行了实现,该算法能够满足 FCD 大数据处理的实时性要求,显著缩短了处理时间,较轮询和 Min-Min 调度等算法有效地提高了并行处理的加速比和效率,为 FCD 大数据快速处理提供了一条有效途径.

参考文献(References)

[1] ISAENKO N, COLOMBARONI C, FUSCOG.Traffic dynamics estimation by using raw floating car data [C]// Proceedings of the 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems. 2017: 704-709.

[2] DE FABRITIIS C, RAGONA R, VALENTI G. Traffic Estimation and prediction based on real time floating car data [C]// 11th International IEEE Conference on Intelligent Transportation Systems. Beijing: IEEE Press, 2008:197-203.

[3] YU Q, RONG J, LIU X M. Floating vehicle data detection system in Beijing[C]//2006 IEEE Intelligent Vehicles Symposium. Tokyo, Japan: IEEE Press, 2006: 320-324.

[4] MESSELODI S, MODENA C M, ZANIN M, et al. Intelligent extended floating car data collection [J]. Expert Systems with Applications, 2009, 36(3): 4213-4227.

[5] SONG G, ZHANG F, LIU J, et al. Floating car data-based method for detecting flooding incident under grade separation bridges in Beijing[J]. IET Intelligent Transport Systems, 2015, 9(8): 817-823.

[6] PFOSER D, TRYFONA N, VOISARD A. Dynamic travel time maps - enabling efficient navigation[C]// 18th International Conference on Scientific and Statistical Database Management. Vienna, Astria: IEEE Press, 2006: 369-378.

[7] MICHAEL JONES; YANFENG GENG; DANIEL NIKOVSKI, et al. Predicting link travel times from floating car data [C]// 16th International IEEE Conference on Intelligent Transportation Systems. Hague, Newtherlands: IEEE Press, 2013: 1756-1763.

[8] ISAENKO N, COLOMBARONI C, FUSCO G. Traffic dynamics estimation by using raw floating car data[C]// 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems. Naples, Italy: IEEE Press, 2017: 704-709.

[9] 陈锋,庞昊, 等. KD-50-I-E 平台的 FCD 并行计算与交通动态诱导[J]. 中国科学技术大学学报, 2009,39(5): 558-560.

[10] DENG Z, BAI Y Q. Floating car data processing model based on Hadoop-GIS tools[C]// Fifth International Conference on Agro-GeoInformatics. Tianjing, China: IEEE Press, 2016: 1-4.

[11] ZHANG D B, SHOU Y F, XU J M. The modeling of big traffic data processing based on cloud computing [C]// 12th World Congress on Intelligent Control and Automation. Guilin, China: IEEE Press, 2016: 2394-2399.

[12] LU W, WANG W J, KIMITA K, et al. Decreasing FCD processing delay by deploying distributed processing system[C]// 6th International Conference on ITS Telecommunications. Chengdu, China: IEEE Press, 2006: 206-209.

[13] ZHANG Z H, JIANG C J, FANG Y. Road situation modeling and parallel algorithm implementation with FCD based on principle curves [C]// Eighth International Conference on High-Performance Computing in Asia-Pacific Region. Beijing, China: IEEE Press, 2005: 6 -11.

[14] CHAUHAN SS, JOSHI R C. QoS guided heuristic algorithms for grid task scheduling. International Journal of Computer Applications, 2010, 2(9): 24-31.

[15] KOKILAVANI T, AMALARETHINAM D I. Load balanced Min-Min algorithm for static meta-task scheduling in grid computing[J]. International Journal of Computer Applications, 2011, 20(2): 43-49.

[16] 朱虹宇,李挺,闫建恩,等. 基于动态负载均衡的分布式任务调度算法研究[J]. 高技术通讯, 2014, 24(12): 1261-1269.

第48卷第9期 中国科学技术大学学报 Vol.48,No.9
2018年9月 JOURNAL OF UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA Sep. 2018

文章编号:0253-2778(2018)09-0723-07

剪枝技术在占优查询中的应用

孙 志,孙雪姣

(烟台大学计算机与控制工程学院,山东烟台 264005)

摘要:用户的偏好在很多情况下可以引导用户的选择,有关偏好查询的问题在关系型数据库中成为越来越重要的问题.在很多应用中,相对于定量偏好,定性偏好能够应用的范围更广.已有的多属性偏好研究中偏好属性都不具有依赖关系,而CP-nets是一种表示具有依赖关系的多属性定性偏好的图模型.目前,对偏好查询的处理主要使用占优查询,通过用户的偏好依次比较两个配置,从而得出满足用户偏好的配置.对配置进行两两比较会造成极大的资源浪费,为了降低其配置的比较次数.提出将剪枝技术应用于占优查询中,通过对翻转序列的路径进行修剪,从而有效地减少数据库搜索的空间.

关键词:条件偏好网;CP-nets导出图;翻转序列;后缀固定;最小变量翻转;向前修剪技术

中图分类号:TP305 **文献标识码**:A doi:10.3969/j.issn.0253-2778.2018.09.006

引用格式:孙志,孙雪姣. 剪枝技术在占优查询中的应用[J]. 中国科学技术大学学报,2018,48(9):723-729.
SUN Zhi, SUN Xuejiao. Application of the pruning technique in dominant query[J]. Journal of University of Science and Technology of China, 2018,48(9):723-729.

Application of the pruning technique in dominant query

SUN Zhi, SUN Xuejiao

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

Abstract: User preferences can influence the user choices in many cases, and the question of preference query becomes an increasingly important issue in relational databases. In many applications, qualitative preferences can be applied more widely than quantitative preferences. In the existing studies of multi-attribute preferences, preference attributes do not have a dependency relationship, but CP-nets(conditional preference networks) is a graph model that represents multi-attribute qualitative preferences with dependencies. At present, the processing of preference queries mainly uses dominance queries and compares the two outcomes one by one to obtain the outcome that satisfies the user preferences. It can be found that comparing the outcomes in pairs causes great waste of resources. Reduce the number of comparisons of its outcomes is examined. The pruning technique is proposed to be applied to dominance queries, and the path of the flipping sequence is pruned so as to effectively reduce the space for database search.

Key words: conditional preference networks(CP-nets); the induced graph of CP-nets; flipping sequence; suffix fixing; least-variable flipping; forward pruning

收稿日期:2018-05-24;**修回日期**:2018-09-18

基金项目:山东省自然科学基金(ZR2014FL009ZR2014FL009),山东省高等学校科技计划项目(OJ14LN23)资助.

作者简介:孙志,男,1993年生,硕士生. 研究方向:数据挖掘.E-mail:sunzhi2016@126.com

通讯作者:孙雪姣,副教授.Email:sunxuejiao6@sina.com

0 引言

偏好在很多情况下可以引导用户的选择.用户的偏好可以是定量偏好,也可以是定性偏好.对于定量偏好,人们提出了用效用函数来描述用户的偏好,基于效用函数将用户的偏好映射到一个数量范围内[1],根据数量的大小对用户的偏好进行排序,但在很多实际问题中无法定量描述用户的偏好.相对于定量偏好,在很多领域中我们更希望用定性的方式来表达用户的偏好.通过定义一个二元关系来表示偏好关系[2],进而对用户的偏好进行排序.

本文使用条件偏好网来指定偏好关系.CP-nets可以通过使用条件偏好语句表示偏好关系,对用户的偏好进行分析从而找出最优集.CP-nets 是一个定性的图形化的描述工具,能够反映偏好之间的依赖关系,在许多实际应用中它的描述表现得更加严密和自然.

目前偏好的应用大多基于 CP-nets 来加以描述[3-9],其中典型的是 Brafman 等[3] 和 Boutilier 等[4]的工作,他们详细描述了 CP-nets 的语法、语义及应用.近年来,对 CP-nets 的子类也进行了研究,如 TCP-nets[10]]和 CI-nets[11],但 CP-nets 上的偏好查询还有待进一步优化.本文的主要创新点在于将剪枝技术应用于占优查询中,在不影响搜索过程的正确性及完整性的情况下,通过应用剪枝技术对搜索树进行修剪,从而减少搜索空间,提高查询效率;并将使用剪枝技术的占优查询与深度优先搜索 DFS 算法进行对比,进一步论证剪枝技术在减少搜索空间、提高查询效率方面的优势.

1 CP-nets 语法与语义

本节首先定义了偏好的相关概念,随后引入了 CP-nets 及其语义.

1.1 偏好的相关定义

定义 1.1[12] 设 $V=\{X_1,X_2,\cdots,X_n\}$ 是决策属性的集合,$\mathrm{dom}(X_i)$ 代表属性 $X_i\in V$ 的有限定义域 $\{x_1^i,x_2^i,\cdots,x_n^i\}$,则决策空间 Ω 是各个属性定义域的笛卡儿积,即 $\Omega=\mathrm{dom}(X_1)\times\cdots\times\mathrm{dom}(X_n)$ 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,则称 o 和 o' 为可交换的配置.

如果决策的两个属性之间有依赖关系,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 $\mathrm{pa}(X_i)$ 来表示.

定义 1.2[12] $\succ$ 是决策空间 Ω 上的二元关系.

(Ⅰ)若 $\succ$ 自反($(\forall o)o\in\Omega\rightarrow o\succ o$)、反对称($(\forall o,o')(o,o'\in\Omega\wedge o\succ o'\wedge o'\neq o\rightarrow o'\cdot o)$)和传递性($(\forall o,o',o'')(o,o',o''\in\Omega\wedge o\succ o'\wedge o'\succ o''\rightarrow o\succ o'')$),即 $\succ$ 是严格偏好排序关系时称 $\succ$ 为 Ω 上的严格偏好关系.

(Ⅱ)若两个配置 $o,o'\in\Omega$ 有 $o\succ o'$ 或 o' 与 o 不可比较,含义是 o' 不比 o 强,表示为 $o\succ_N o'$,即 $\succ_N$ 为 CP-nets 所能表达的偏好集合.

二元关系 $\succ$ 反映了决策者对两个配置 o,o 的偏好强弱关系,即 $o\succ o'$ 表示决策者对 o 的偏好优于 o' 的偏好

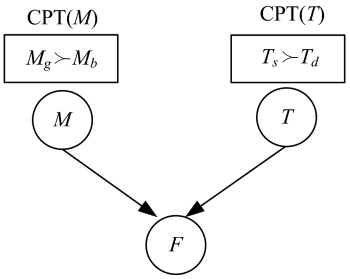
1.2 条件偏好图—CP-nets

定义 1.3[12] $\mathrm{CPT}(X_i)$ 为属性 X_i 的条件偏好表,它表示属性 X_i 在 $\mathrm{pa}(X_i)$ 的不同取值下决策者对 $\mathrm{dom}(X_i)$ 集合的一个偏好排序,即在 $\mathrm{pa}(X_i)$ 的不同取值下,对属性 X_i 的所有取值的偏好排序也不一样.此时,由决策者提供的 $\mathrm{pa}(X_i)$ 的一种取值成为属性 X_i 的一个决定条件.那么,在所有的决定条件下,决策者对属性 X_i 的取值的偏好排序构成了属性 X_i 的条件偏好表 $\mathrm{CPT}(X_i)$.

定义 1.4[12] CP-nets 是一个有向图 $N=\langle V,CE\rangle$,其中 V 是顶点集(决策属性集),每一个顶点 X_i 都有一个条件偏好表 $\mathrm{CPT}(X_i)$ 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着终点取值的偏好.

例 1.1 某人观看电影主要考虑 3 个方面:心情、影片类型和零食,分别用 M, T 和 F 代表.心情无条件喜欢好心情而不是坏心情;影片类型无条件喜欢动作片而不是剧情片;零食吃否则取决于心情和所观看的影片类型.如果心情好且观看的是剧情片的话,则他喜欢吃零食;如果心情不好且观看的是剧情片的话,则他不喜欢吃零食;如果心情好且观看的是动作片的话,则他不喜欢吃零食;如果心情不好且观看的是动作片的话,则他喜欢吃零食.

该例中的 CP-nets $N=\langle V,CE\rangle$ 如图 1 所示,其中 $V=\{M,T,F\}$, $\mathrm{dom}(M)=\{M_g,M_b\}$, $\mathrm{dom}(T)=\{T_s,T_d\}$, $\mathrm{dom}(F)=\{F_y,F_n\}$, $CE=\{\langle M,F\rangle,\langle T,F\rangle\}$, 各顶点的条件偏好表在图 1 中以表格形式给出.



CPT(F)

M_gT_d	$F_y \succ F_n$
M_bT_d	$F_n \succ F_y$
M_gT_s	$F_n \succ F_y$
M_bT_s	$F_y \succ F_n$

图 1　观看电影的 CP-nets

Fig.1　CP-nets for "watching movies"

定义 1.5[12]　设 $N=\langle V,CE\rangle$ 是一个 CP-nets，则有向图 $G=\langle \Omega,IE\rangle$ 是 N 的导出图，其中 Ω 是顶点集，IE 是可交换的配置所构成的有向边集.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 $o_j \succ o_i$.

例 1.1 中 CP-nets 对应的导出图如图 2 所示.对于导出图 $G=\langle \Omega,IE\rangle$，$IE$ 是可交换的配置所构成的有向边集.对于二值 CP-nets，每一条有向边 $\langle o_i,o_j\rangle$ 表示将配置 o_i 中某个属性值翻转得到配置 o_j.

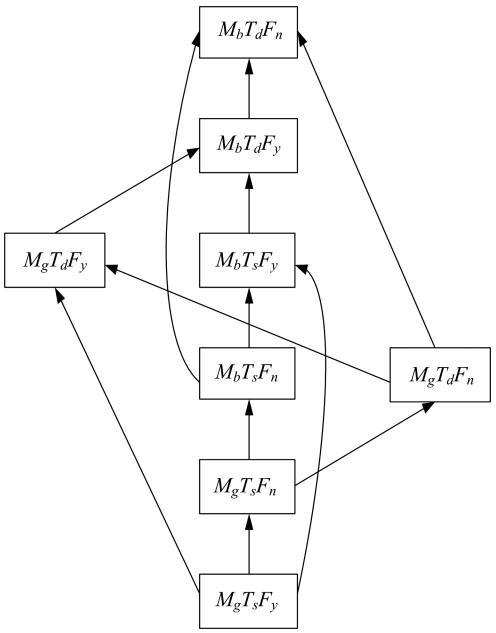


图 2　CP-nets 导出图

Fig.2　The induced graph of CP-nets

1.3　CP-nets 的语义

基于 CP-nets 的偏好遵循 ceteris paribus，即对于一个配置 $o\in\Omega$ 来说，主要关心除某个属性 X_i 的值不同外，其他属性值都相同，决策者对属性 X_i 不同取值的一种偏好断言.利用 ceteris paribus 语义，容易对可交换的两个配置给出偏好断言.因为可交换的两个配置的对比本质上是一个属性值的对比.

定义 1.6[12]　设 $X,Y,Z\subset V$，且 $X\cup Y\cup Z=V$，$X\cap Y=\varnothing$，$X\cap Z=\varnothing$，$Y\cap Z=\varnothing$，对于 Z 的一个赋值 z 和任意的 $x_1,x_2\in \mathrm{dom}(X)$，任意的 $y_1,y_2\in \mathrm{dom}(Y)$，当 $x_1\,y_1z \succ x_2\,y_1z$ 时，必有 $x_1\,y_2z \succ x_2\,y_2z$，则称属性 X 与属性 Y 在条件 Z 的一个赋值 z 下条件偏好无关.

对于例 1 中的配置，有关系：$B_i\,T_d\,C_i \succ B_d\,T_d\,C_i$，$B_i\,T_l\,C_i \succ B_d\,T_l\,C_i$，即属性 B 与属性 T 在 C 的一个赋值 C_i 下条件偏好无关.

2　占优查询及翻转序列

给定一个 CP-net N，配置 o 和 o' 之间的关系一定为以下 3 种：$N\models o\succ o'$，$N\models o'\succ o$，或者 $N\not\models o\succ o'$ 及 $N\not\models o'\succ o$.第 3 种关系是指 N 中没有足够的信息表明一个配置优于另一个配置.

定义 2.1　给定一个 CP-net N 以及配置 o 和 o'.如果 $o\succ o'$，且 $N\not\models o'\succ o$，则称 o 占优于 o'.判断 $o\succ o'$ 是否成立的查询称为占优查询.

定义 2.2[13]　设 $G=\langle \Omega,IE\rangle$ 是 CP-net N 的导出图，对于配置 $o,o'\in\Omega$，若存在一条路径连接顶点 o 和 o'，则称 o 占优于 o'，记作 $o\succ o'$.

由于 IE 是可交换的配置所构成的有向边集，故称为翻转关系.对于二值 CP-nets，若顶点 o 和 o' 存在一条路径连接，则可将 o' 的一个属性值翻转(取两个属性值的另一个)得到 o.$o\succ o'$ 说明配置 o 比配置 o' 优，即在 CP-net 导出图上存在一个翻转序列，使得该序列连接顶点 o' 和 o.

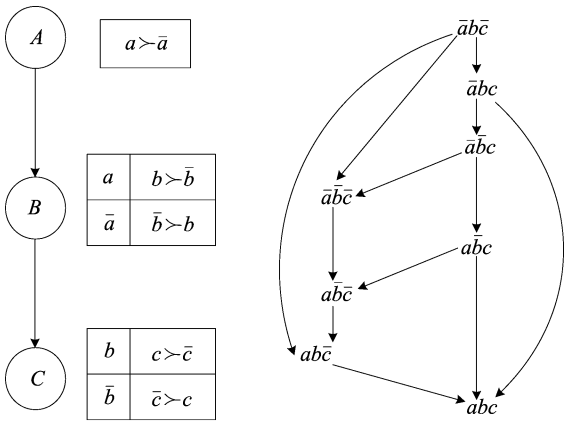
综上可知，CP-nets 的 ceteris paribus 语义允许使用变量 X 的 CPT(X) 中的信息来改变或翻转配置中 X 的值，以得到改善的或恶化的配置.从一个配置到另一个配置的改善翻转序列表明，在满足该 CP-nets 的所有排序中，一个配置是否偏好于另一个配置.在更确切地定义这个概念之前，我们用一个例子来说明.

例 2.1　考虑图 3 中的 CP-net.满足该 CP-net 的排序如下：

$abc \succ ab\bar{c} \succ a\bar{b}\bar{c} \succ a\bar{b}c \succ \bar{a}\bar{b}\bar{c} \succ \bar{a}\bar{b}c \succ \bar{a}bc \succ \bar{a}b\bar{c}$,

$abc \succ ab\bar{c} \succ a\bar{b}\bar{c} \succ \bar{a}\bar{b}\bar{c} \succ a\bar{b}c \succ \bar{a}\bar{b}c \succ \bar{a}bc \succ \bar{a}b\bar{c}$.

由上可知,我们无法根据现有信息确定配置 $\bar{a}\bar{b}\bar{c}$ 和 $a\bar{b}c$ 的偏好关系.如果我们从每个配置链中删除 $\bar{a}\bar{b}\bar{c}$ 或 $a\bar{b}c$,那么根据其 CPT(X) 中的偏好信息,我们可以通过翻转一个变量的值从一个配置移动到另一个配置.例如,为了从这些序列中的第一个配置 (abc) 移动到第二个配置 ($ab\bar{c}$),我们根据在给定 b 的情况下有 $c \succ c'$ 来表明第一个配置优于第二个配置,即在给定其父亲 B 的实例化 b 的情况下,我们将 C 翻转为一个次优值.



(a) CP-nets and its CPT (b) Induced graph

图 3 CP-nets 及其导出图

Fig.3 CP-nets and its induced graph

通过上述分析,我们得出对于任意两个配置 o 和 o',从 o' 到 o 的每个改进翻转序列与 CP-nets 导出图中从节点 o' 到节点 o 的有向路径相对应.例如,考虑图 4(a)中的 CP-nets,从配置 $\bar{a}\bar{b}c$ 到配置 abc 有 4 条可选的翻转序列,与图 4(b)导出图中的这些配置之间的 4 条有向路径相对应,即

$\bar{a}\bar{b}c \to a\bar{b}c \to abc$,

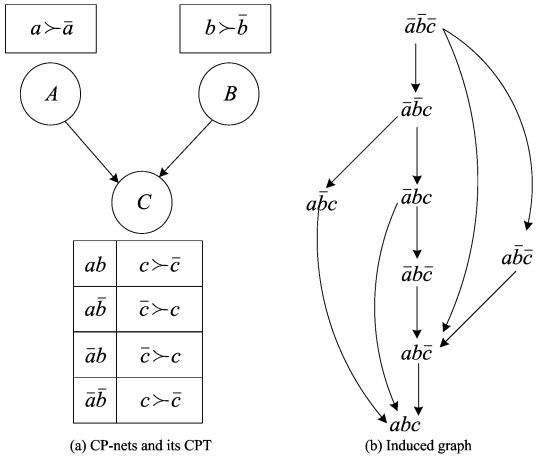
$\bar{a}\bar{b}c \to a\bar{b}c \to a\bar{b}\bar{c} \to ab\bar{c} \to abc$,

$\bar{a}\bar{b}c \to \bar{a}bc \to abc$,

$\bar{a}\bar{b}c \to \bar{a}bc \to \bar{a}b\bar{c} \to ab\bar{c} \to abc$;

因此,$abc \succ \bar{a}\bar{b}c$ 是这个 CP－net 的结果.由于从 $\bar{a}bc$ 到 $a\bar{b}\bar{c}$ 没有翻转序列(有向路径),根据现有信息无法对这两个配置进行比较.

这些例子表明,这种翻转序列的构建可以用来证明偏好关系.



(a) CP-nets and its CPT (b) Induced graph

图 4 CP-nets 及其导出图

Fig.4 CP-nets and its induced graph

定义 2.3 令 N 为变量集 V 的 CP-net, $X_i \in V$, $U = \mathrm{Pa}(X_i)$, $Y = V\text{-}(U \cup \{X_i\})$. 令 $uxy \in \mathrm{Asst}(V)$ 为任一配置,其中 $x \in \mathrm{dom}(X_i)$, $u \in \mathrm{Asst}(V)$, $y \in \mathrm{Asst}(Y)$, 变量 X_i 的配置 uxy 的改进翻转是使得 $x' \succ_{i_u} x$ 成立的任一配置 uxy. 如果在给定 u 的情况下 x 是 X_i 的最优值,则 X_i 不存在改进翻转.

如果对于任意 $i < k$, 都有 o_{i+1} 是某个变量 o_i 的改进翻转,则关于 N 的改进翻转序列是指配置 $o_1, \cdots, o_k$ 的任意序列.从一个配置 o 到另一个配置 o' 的改进翻转序列是 $o_1, \cdots, o_k$ 的任一改进序列,其中 $o_1 = o$, $o_k = o'$.

占优查询对二值 CP-nets 的处理需要大量的翻转序列,文献[4]给出了不同图结构的时间复杂度分析如下.

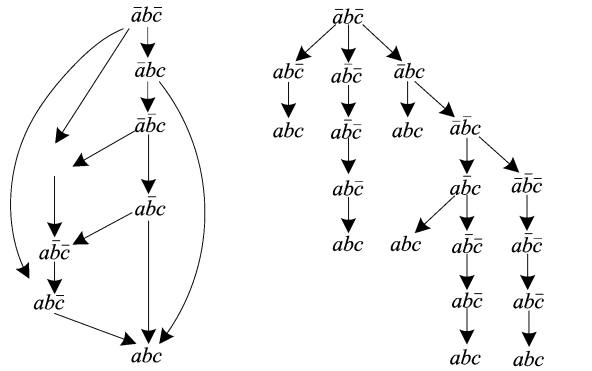
当 CP-nets 图结构为有向树时,其时间复杂度为 $O(n^2)$; 当 CP-nets 图结构为 polytree(即导出图是无向非循环的)时,其时间复杂度为多项式时间复杂度 P;当 CP-nets 图结构为单连接有向图(即任一对节点之间最多存在一条有向路径)时,其时间复杂度为 NP 完备.

由此可知,使用占优查询对二值 CP-nets 翻转序列的处理存在困难.本文提出使用剪枝技术对翻转序列进行优化,减少需要的翻转序列数量,进而提高查询效率.

3 占优查询的剪枝技术

本节讨论搜索翻转序列和几个规则,在不影响搜索正确性或过程完整性的前提下对搜索树进行修剪.这些规则可应用于改进翻转序列中.

给定一个 CP-net$N$ 和一个配置 o，我们定义 o 的改进搜索树 $T(o)$ 如下：$T(o)$ 的根为 o，$T(o)$ 中每个节点 o' 的子代都是那些可以通过一个从 o' 出发的改进翻转可达的配置.容易得出，$T(o)$ 中的根路径对应于配置 o 的改进翻转序列；反之亦然.例如，考虑图 5(a)所示的导出图，其由图 3 中的 CP-net 得出.图 5(b)描述了关于该偏好图的改进搜索树 $T(\bar{a}b\bar{c})$. 显然，可以把每个占优查询 $N\models o\succ o'$ 看作是从根节点 o' 开始到配置 o 的搜索树 $T(o')$. 在上面的例子中，给定占优查询 $N\models a\bar{b}\bar{c}\succ\bar{a}b\bar{c}$，由图 5(c)中所示的虚线路径可得到配置 $a\bar{b}\bar{c}$.



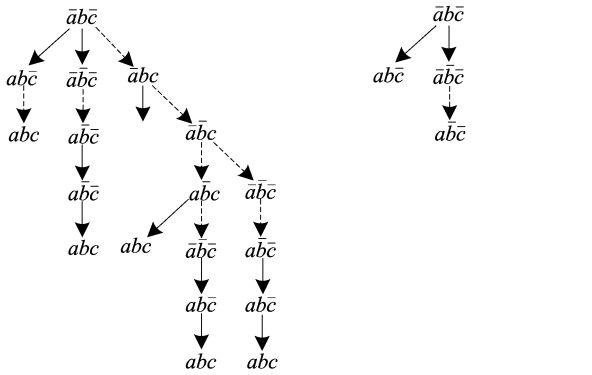


图 5　使用后缀固定技术修剪搜索树 $T(\bar{a}b\bar{c})$

Fig.5　Pruning the search tree $T(\bar{a}b\bar{c})$ by suffix fixing

3.1　后缀固定规则

后缀固定规则是指在不影响搜索完整性的情况下对搜索树 $T(o)$ 进行指定路径的修剪.令 N 为变量集 $V=\{X_1,\cdots,X_n\}$ 的 CP-net，根据与 N 一致的拓扑排序进行编号.对任意的 $r\geq 1$，我们把配置 $o\in\mathrm{Asst}(V)$ 的第 r 个后缀定义为配置值 $o[X_r]o[X_{r+1}]\cdots o[X_n]$ 的子集，其中配置的第 r 个后缀取决于使用变量的拓扑排序.当对所有的 $r\leqslant j\leqslant n$，都有 $o[X_j]=o'[X_j]$，则称配置 o 为 o' 的第 r 个后缀匹配.

令 o^* 是改进搜索树 $T(o')$ 中的一个节点，并且令 o 为该搜索的目标配置，即试图找到证明 $N\models o\succ o'$ 的翻转序列.根据后缀固定规则，如果 o^* 和 o 的第 r 个后缀匹配，则不需搜索那些与配置 o 的第 r 个后缀不匹配的配置.可以得出，如果在搜索树 $T(o^*)$ 中有从根 o^* 到 o 的路径，则存在一个从根 o^* 到 o 的路径，使得该路径上的每个配置对 $X_r,\cdots,X_n$ 分配相同的值.如图 5(c)中有向路径：$\bar{a}b\bar{c}\to\bar{a}\bar{b}\bar{c}\to a\bar{b}\bar{c}$，其中每个配置对 C 分配相同的值 $\bar{c}$. 假设给定查询 $N\models o\succ o'$，在 $T(o')$ 中存在一条路径到配置 o^*. 由于以 o^* 为根的 $T(o')$ 的子树是 $T(o^*)$，因此如果在改进搜索树中存在从 o' 到 o 的路径通过 o^*，则 CP-nets 导出图中存在一条从 o' 到 o 的路径通过 o^*，其中 o^* 的后缀保留在从 o^* 到 o 的子路径上.从而得出，如果使用后缀规则进行修剪，则改进搜索树保持其完整性.

综上所述，后缀固定规则有效地修剪节点 o^* 下的搜索树，仅保留与目标配置 o 的后缀值相同的路径.例如，对于图 5(a)中的查询 $N\models a\bar{b}\bar{c}\succ\bar{a}b\bar{c}$，如果使用 DFS 算法进行查询，需以配置 $\bar{a}b\bar{c}$ 为出发点依次从其出发搜索它的每个邻接点，若该邻接点未曾访问过，则以该邻接点为新的出发点继续进行深度优先遍历，直至发现目标 $a\bar{b}\bar{c}$ 或图中所有和出发点 $\bar{a}b\bar{c}$ 有路径相通的顶点均已被访问为止.由图 5(a)知，满足该查询的路径共有 3 条，分别为 $\bar{a}b\bar{c}\to\bar{a}\bar{b}\bar{c}\to a\bar{b}\bar{c}$、$\bar{a}b\bar{c}\to\bar{a}bc\to\bar{a}\bar{b}\bar{c}\to a\bar{b}\bar{c}$、$\bar{a}b\bar{c}\to\bar{a}bc\to\bar{a}\bar{b}c\to a\bar{b}c\to a\bar{b}\bar{c}$，共需进行 10 次比较；如果使用后缀固定规则及变量排序 A,B,C 对改进搜索树 $T(\bar{a}b\bar{c})$ 进行修剪，如图 5(c)中虚线所示均为满足查询 $N\models a\bar{b}\bar{c}\succ\bar{a}b\bar{c}$ 的有向路径，根据上文使用后缀规则进行修剪，改进搜索树保持其完整性可得出如图 5(d)中所示的有向路径 $\bar{a}b\bar{c}\to\bar{a}\bar{b}\bar{c}\to a\bar{b}\bar{c}$，从而发现只需进行 2 次比较即可得出满足用户偏好的配置.由此可知，通过后缀固定规则对搜索树进行修建可有效地减少搜索树的大小，即减少搜索路径，进而减少配置间的比较次数，提高查询效率.

3.2　最小翻转变量

最小翻转变量规则是后缀固定规则的扩展.假设有一个 CP-net N 以及查询 $N\models o\succ o'$. 令 o^* 为其中的一个配置，对于任何变量 X_j，令 u 表示 o^* 中 $U=\mathrm{pa}(X_j)$ 的实例.如果存在 $x\in\mathrm{dom}(X_j)$ 使得 $x\succ_{j_u}o^*[X_j]$ 成立，并且 N 中 X_j 的子代都没