计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.200800082

基于不相关属性集合的属性探索算法

沈夏炯[1,2]　杨继勇[1,2]　张　磊[1,2,3]
1 河南大学河南省大数据分析与处理重点实验室　河南 开封 475000
2 河南大学计算机与信息工程学院　河南 开封 475000
3 河南大学数据与知识工程研究所　河南 开封 475000
(shenxj@henu.edu.cn)

摘　要　作为形式概念分析理论中的一个重要工具,属性探索算法能够以问题为导向,交互式地逐步发现系统知识,在知识的发现和获取中居于核心地位。但是,当形式背景的规模较大时,属性探索算法的计算过程过于耗时,严重制约了算法在当前大数据时代的推广与应用。耗时瓶颈主要存在于"寻找下一个与专家交互的问题"这一环节,传统算法在此过程中存在大量冗余计算。针对这个问题,在分析伪内涵和内涵与蕴涵集合的内在逻辑关系的基础上,提出并证明了3个定理,根据定理给出了一种基于不相关属性集合的属性探索算法,该算法在计算伪内涵与内涵的过程中,借助提出的定理,跳过违反该逻辑关系的属性集合是否为伪内涵或者内涵的判断过程,减小了算法的搜索空间,从而降低了算法的时间复杂度。所提算法最好的时间复杂度为$O(mn^2P^2)$,最坏的时间复杂度为$O(mn^3P^2)$。实验结果表明,与传统算法相比,该算法具有较为明显的时间性能优势。

关键词:形式概念分析;伪内涵;关联规则;属性探索;概念格;知识发现
中图法分类号　TP301

Attribute Exploration Algorithm Based on Unrelated Attribute Set

SHEN Xia-jiong[1,2], YANG Ji-yong[1,2] and ZHANG Lei[1,2,3]
1 Henan Key Laboratory of Big Data Analysis and Processing, Henan University, Kaifeng, Henan 475000, China
2 College of Computer and Information Engineering Henan University, Kaifeng, Henan 475000, China
3 Institute of Data and Knowledge Engineering, Henan University, Kaifeng, Henan 475000, China

Abstract　As an important tool in the theory of formal concept analysis, the attribute exploration algorithm is problem-oriented and can interactively discover system knowledge step by step, which plays a central role in knowledge discovery and acquisition. However, if the size of formal context is large, the calculation process of attribute exploration algorithm will spend too much time to restrict seriously the promotion and application of the algorithm in the current era of big data. The bottleneck of time-consuming mainly lies in "finding the next problem to interact with experts", traditional algorithms have a lot of redundant computation in this process. Aiming at this problem, three theorems are put forward and proved based on analyzing the logic relation between pseudo-intent, intent and implication set. According to these theorems, an attribute exploration algorithm based on an unrelated collection is given. During pseudo-intent and intent calculation, this algorithm, by means of the proposed theorems, can skip the process of determining whether or not an attribute set that violates the logical relationship is a pseudo-intent or intent, so as to reduce the search space and time complexity of the algorithm. The best time is $O(mn^2P^2)$, the worst time is $O(mn^3P^2)$. The experimental results show that the proposed algorithm has an obvious time performance advantage compared with the traditional algorithm.

Keywords　Formal concept analysis, Pseudo-intent, Association rules, Attribute exploration, Concept lattice, Knowledge discovery

到稿日期:2020-08-13　返修日期:2020-11-21　本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:国家自然科学基金(61701170);河南省科技厅科技攻关计划基金(202102310340);河南省高等学校青年骨干教师培养计划项目(2019GGJS040,2020GGJS027);河南省高等学校重点科研项目(21A110005)
This work was supported by the National Natural Science Foundation of China(61701170), Scientific and Technological Project of Henan Province(202102310340), Foundation of University Young Key Teacher of Henan Province(2019GGJS040, 2020GGJS027) and Key Scientific Research Projects of Colleges and Universities in Henan Province(21A110005).
通信作者:张磊(zhanglei@henu.edu.cn)

1 引言

形式概念分析(Formal Concept Analysis,FCA)[1]把形式背景中具有共同属性的对象抽取出来,加以概括后称为形式概念,进而通过形式概念之间的泛化-特化关系来构造知识的层次关系。概念格将形式概念通过 Hasse 图可视化,直观地描述了对象(样本)与属性(特征)之间的关联,因此概念格理论与方法是形式概念分析研究中的基本内容[2]。

形式概念分析被认为是数据分析和知识描述的有利工具,目前已被广泛应用于数据分析[3]、知识发现[4]、规则提取[5]、概念认知学习[6-7]等领域[8]。

属性探索算法[9]能够通过交互式询问领域专家一系列问题的方式,来计算形式背景中属性之间的蕴涵关系。这种蕴涵关系能够表示事物的内在逻辑规律,是知识表示的一种重要形式。与其他基于对已有数据进行统计或归纳方式的知识获取方法不同,属性探索算法能够通过询问的方式主动探索未知蕴涵关系。其探索过程可概括为:首先,根据形式背景中的属性集合,提出属性与属性间蕴涵关系是否成立的问题;其次,由专家判断这些问题是否成立;然后,属性探索算法根据专家不同的回答,计算下一个需要交互的属性与属性间的蕴涵关系式;最后,算法循环产生问题,从而获取到领域专家知识中属性之间的所有蕴涵关系。

例如,在信息系统访问控制需求捕获的问题中,我们以访问控制权限为属性,以用户为对象建立该系统的访问控制形式背景。因算法对访问控制授权的背景知识未知,算法初始时的访问控制形式背景为空,之后算法根据权限集合提出了类似于"实验室的所有人都可以打开实验室的门?""可以打开实验室门的人是不是都可以打开保险柜的门?"的蕴涵关系问题。这些问题由交互的领域专家给出一个肯定的答案,或者提出一个反例反驳它。如果专家给出肯定的答案,则记录这个蕴涵关系式,并计算出下一个待专家回答的蕴涵关系式。如果专家给出了一个反例,则将反例加入算法的形式背景中,并在加入反例后的形式背景中重新计算当前的蕴涵关系式。算法如此反复迭代直至最终获取到领域专家关于访问控制授权的知识背景,以及权限(属性)间的蕴涵关系。这种通过主动询问专家的方式来探索未知属性蕴涵关系的特点,使得属性探索算法在知识发现和获取中具有核心地位,被认为是一种重要的知识获取方法。

目前,属性探索算法已经被应用于多个领域。Borchmann[10]从形式概念分析出发,为属性探索提出了一个通用的描述框架,将属性探索的变体看作通用框架下的一个实例。Borchmann[11]提出了基于置信度概念的属性探索,为在可能有错误的形式背景中进行属性探索,提供了一个可行性方案。Glodeanu[12]提出了基于已有背景知识的模糊属性探索算法,该算法允许领域专家给出模糊的回答。Obiedkov 等[13]以属性探索交互式地构建基于网络结构的访问控制模型。Jäschke 等[14]将属性探索应用于 Web 查询,提出了一种基于属性探索的网络信息检索方法,提高了查询效率。Obiedkov 等[15]提出了协作的概念探索,帮助人们构建事物本体。Hanika 等[16]提出了一种专家对概念集进行协作获取知识的方法。Codocedo 等[17]使用属性探索算法对上下文进行抽样,提出了一种计算上下文模式结构的方法,降低了计算高维对象结构的复杂度。

然而,由于属性探索算法具有属性蕴涵计算完备性的特点,因此其计算过程过于耗时,无法满足当前大数据时代中对海量数据知识获取的需求,限制了算法的进一步应用。如何降低计算的时间开销,成为了当前研究的一个重要问题。文献[18-19]为属性探索算法设置了一个固有前提,以提高算法效率,但是这种方案在提高算法效率的同时,也使得该算法具有一定的局限性。

Kriegel[20]提出了一种并行的属性探索算法,以增加单位时间计算能力的方式来降低属性探索算法的耗时。并行计算的方案缩短了属性探索算法的整体耗时,但是无助于改进单个节点中串行算法的时间复杂度。与前述方案不同,本文着眼于寻找并规避算法内在的冗余计算过程,以有效降低算法的时间开销。

属性探索算法耗时瓶颈的关键在于"寻找下一个与专家交互的问题",传统属性探索算法以计算蕴涵伪壳的方式寻找下一个交互问题。研究发现,此过程中存在规避冗余计算的可能。例如,Zhao 等[21]运用属性集合与蕴涵集合相关性定义对属性探索算法进行改进,在一定程度上改进了算法的效率。但是该算法在寻找下一个交互问题时,需要逐个遍历属性集合的所有组合方式,当属性较多时算法的搜索空间较大,耗时仍较为明显。

本文发现,在属性探索过程中存在一类与主基不相关的集合。这些集合包含主基中某个蕴涵式的前件,不包含这个蕴涵式的后件,而且主基与内涵集合都不含有这类属性集合。

对此,本文提出了一种 AEUS(Attribute Exploration of Unrelated Set)算法,借助属性集合与主基不相关的关系,跳过与主基不相关的属性集合是否为下一个属性探索问题的判断过程,减小寻找下一个交互问题的搜索空间,从而降低算法的时间复杂度。

2 基础知识

下面给出本文所用到的相关定义[1,5,21]。

定义 1 一个形式背景 $K=(U,M,I)$ 由两个集合 U 和 M 以及 U 与 M 间的关系 I 组成,U 的元素称为对象,M 的元素称为属性。$(u,m)\in I$ 或者 (uIm) 表示对象 u 拥有属性 m。我们用 $(u,m)\notin I$ 表示对象 u 不拥有属性 m。

定义 2 设 $K=(U,M,I)$ 是一个形式背景,若 $A\subseteq U$,$B\subseteq M$,令:

$$f(A)=\{m\in M\mid \forall u\in A,(u,m)\in I\}$$

及

$$g(B)=\{u\in U\mid \forall m\in B,(u,m)\in I\}$$

如果 A,B 满足 $f(A)=B,g(B)=A$,则称二元组 (A,B) 是一个概念。A 是概念 (A,B) 的外延,B 是概念 (A,B) 的内涵。

形式背景 $K=(U,M,I)$ 上的概念有以下基本性质($\forall A,A_1,A_2\subseteq U,\forall B,B_1,B_2\subseteq M$)。

性质 1 $A_1\subseteq A_2\Rightarrow f(A_2)\subseteq f(A_1)$;$B_1\subseteq B_2\Rightarrow g(B_2)\subseteq g(B_1)$;$A\subseteq g(f(A))$;$B\subseteq f(g(B))$;若 $B=f(g(B))$,则 B 是

形式背景 K 上的一个内涵。

定义 3 设 $K=(U,M,I)$是一个形式背景，$Y\subseteq M$，满足：

(1)$Y\neq f(g(Y))$(即 $Y\subset f(g(Y))$)；

(2)对于每一个伪内涵$Y_1\subset Y$ 都有 $f(g(Y_1))\subseteq Y$，则称 Y 是一个伪内涵。

定义 4 设 $K=(U,M,I)$是一个形式背景，$Y_1,Y_2\subseteq M$。若 $g(Y_1)\subseteq g(Y_2)$，则称在 K 中Y_2值依赖于Y_1，记作$Y_1->Y_2$，也称蕴涵式$Y_1->Y_2$在 K 中成立。

定义 5 设 $K=(U,M,I)$是一个形式背景，则称值依赖集合$\{B->f(g(B))-B|B$ 是 K 的伪内涵$\}$是 K 的主基。

定义 6 给定形式背景 $K=(U,M,I)$，蕴涵式集合 $J(K)$，蕴涵式 $C->D\in J(K)$。若属性集合 $T\subseteq M$ 当且仅当 $C\nsubseteq T$ 或 $D\subseteq T$ 时，则称 T 与 $C->D$ 相关。若 T 与 $J(K)$中所有的蕴涵式都相关，则称 T 与 $J(K)$相关。

根据概念格的值依赖理论，主基可以产生在形式背景中成立的全部值依赖，即属性的蕴涵关系。由定义 5 可知，只要找出全部伪内涵即可得到形式背景的主基。定义 6 中属性集合与蕴涵式集合的相关性判断可以用于伪内涵的计算。

定义 7 设 $K=(U,M,I)$是一个形式背景，$M=\{m_1,m_2,\cdots,m_n\}$，M中的属性满足基本线性序关系($m_1<m_2<\cdots<m_n$)，则对于任意的$Y_1,Y_2\subseteq M$ 当且仅当存在$m_i\in Y_2-Y_1$且 $Y_1\cap\{m_1,\cdots,m_{i-1}\}=Y_2\cap\{m_1,\cdots,m_{i-1}\}$时，称属性集合$Y_1$的字典序小于属性集合$Y_2$的字典序，记作$Y_1<Y_2$。

定义 7 描述的属性集合字典序关系$<$是2^M的一个线性序关系，可按照该字典序的关系逐个产生所有的属性集合，并逐个测试该属性集合是否是伪内涵或者内涵。

3 属性探索算法的理论研究与改进

属性探索算法以主动提出问题的方式与领域专家交互，通过字典序遍历属性集合，并测试该集合是否是伪内涵或者内涵。利用为伪内涵的属性集合产生蕴涵式，从而构建形式背景的主基，获取相关的背景知识。由于该字典序是所有属性幂集上的一个线性序，因此保证了属性探索算法的完备性，但当属性数目较多时，算法的耗时很长。

上述过程中耗时的关键在于，通过遍历的方式测试属性集是否为伪内涵或者内涵的过程存在大量冗余计算。本文发现，如果属性集合包含主基中某个蕴涵式的前件，但是不包含后件，那么这个属性集合就不可能是内涵或者伪内涵，这有助于跳过一些不必要的计算过程。

3.1 理论依据

为方便阐述，我们首先做如下定义。

定义 8 给定形式背景 $K=(U,M,I)$，属性集合 $B,D\subseteq M$，且 $B<D$。若集合 $T=\{C|B<C<D,C\subseteq M\}$，则称 T 是属性集合 B 与属性集合 D 在序$<$上的开区间，记为$\langle B,D\rangle$。

定义 9 给定形式背景 $K=(U,M,I)$，属性集合 $B,B'\subseteq M$。若 $B<B'$且区间$\langle B,B'\rangle$为空集，则称B'仅大于 B，记为 $B'>B$。

定义 10 给定形式背景 $K=(U,M,I)$，属性集合 $B,N\subseteq M$，$B<N$ 且 $N\supsetneq B$，若对于任意的属性集合 $T\in\langle B,N\rangle$，都有 $T\supset B$，则称 N 非平凡仅大于 B，记为 $B\lesssim N$。

定义 11 给定形式背景 $K=(U,M,I)$与 K 上的主基 $J(K)$，蕴涵式 $C->D\in J(K)$。若属性集合 $T\subseteq M$ 当且仅当 $C\subseteq T$ 且 $D\nsubseteq T$ 时，称 T 与蕴涵式 $C->D$ 不相关。若 T 与主基 $J(K)$中任意一个蕴涵式不相关，则称 T 与 $J(K)$不相关。

基于上述定义，我们有如下发现，可作为进一步改进属性探索算法的理论依据。

定理 1 给定形式背景 $K=(U,M,I)$与 K 上的主基 $J(K)$，任意蕴涵式 $C->D\in J(K)$。若属性集合 T 与$C->D$ 不相关，则在 K 中 T 既不是内涵也不是伪内涵。

证明：先证明 T 不是内涵。由于 T 与$C->D$ 不相关，因此由定义 11 可知 $C\subseteq T$ 且 $D\nsubseteq T$。由性质 1 可知 $T\subseteq f(g(T))$，所以 $C\subseteq T\subseteq f(g(T))$。由 $C\subseteq T$ 和性质 1 可知，$g(T)\subseteq g(C)$，$f(g(C))\subseteq f(g(T))$。将 $f(g(C))\subseteq f(g(T))$两端同时减 C 得到 $f(g(C))-C\subseteq f(g(T))-C$。又因为 $C->D\in J(K)$，所以 $f(g(C))-C=D$。因为 $D\nsubseteq T$，所以 $f(g(T))-C\nsubseteq T$，将式子两端同时加上集合 C 得 $f(g(T))\nsubseteq T\cup C$，因为 $C\subseteq T$，所以 $f(g(T))\nsubseteq T$。由性质 1 可知 T 不是内涵。

再证明 T 不是伪内涵。由上述证明知 $f(g(T))\neq T$，即 T 满足定义 3 的条件(1)。

下面利用反证法说明 T 不满足定义 3 的条件(2)。假设 T 满足定义 3 的条件(2)，则任意伪内涵$Y_1\subset T$ 都必须满足 $f(g(Y_1))\subseteq T$。由于 $C->D\in J(K)$，因此 C 在 K 中是一个伪内涵。又因为 T 与$C->D$ 不相关，由定义 11 可知 $C\subseteq T$ 且 $D\nsubseteq T$。根据定义 5 有 $f(g(C))-C=D\nsubseteq T$，所以 $f(g(C))\nsubseteq T$。即，存在一个伪内涵 $C\subset T$ 不满足 $f(g(C))\subseteq T$，与假设命题矛盾。因此，在 K 中 T 不是伪内涵。证毕。

定理 1 表明，如果属性集合与主基中任意一个蕴涵式不相关(即包含蕴涵式前件，但是不包含这个蕴涵式后件)，那么这个属性集合既不是内涵也不是伪内涵。因为在属性探索中只考虑内涵或者伪内涵的属性集合，所以满足定理 1 的属性集合可以忽略不计。

引理 1 设 $K=(U,M,I)$是一个形式背景，对于任意的 $Y_1,Y_2\subseteq M$，若$Y_1<Y_2$，则$Y_2\nsubseteq Y_1$。

证明：因为$Y_1<Y_2$，由定义 7 可知，存在$m_i\in Y_2-Y_1$且 $Y_1\cap\{m_1,\cdots,m_{i-1}\}=Y_2\cap\{m_1,\cdots,m_{i-1}\}$。若$Y_2\subseteq Y_1$，则$m_i=Y_2-Y_1=\varnothing$，即不存在这样的$m_i$，与$Y_1<Y_2$矛盾，所以$Y_2\nsubseteq Y_1$。

定理 2 给定形式背景 $K=(U,M,I)$与 K 上的主基 $J(K)$，对于任意蕴涵式 $B->f((g(B))-B\in J(K)$，若存在 $B',N\subseteq M$，满足$B'>B$ 且B'与 $J(K)$不相关，$B\lesssim N$ 且 N 与 $J(K)$相关，则在区间$\langle B,\min(f((g(B)),N)\rangle$内，既不存在内涵也不存在伪内涵。

证明：因为B'与 $J(K)$不相关，由定理 1 可知B'既不是内涵也不是伪内涵。

(1)设 $f((g(B))>N$，由定义 10 可知，对于任意的属性集合 $C\in\langle B,N\rangle$，都满足 $B\subseteq C$。因为 $C<N<f(g(B))$，所以由引理 1 可知 $f(g(B))\nsubseteq C$，即 C 满足定义 11 的题设条件，所以由定理 1 可知 C 既不是内涵也不是伪内涵。由于 C 是

区间$\langle B,N\rangle$内任意的属性集合，因此在区间$\langle B,N\rangle$内，既不存在内涵也不存在伪内涵。

(2)设$N>f((g(B))$，由区间定义可知$\langle B,f(g(B))\rangle\subset\langle B,N\rangle$。由定义10可知，对于任意的属性集合$C\in\langle B,f(g(B))\rangle$，都满足$B\subseteq C$。因为$C<f(g(B))$，所以由引理1可知$f(g(B))\not\subseteq C$，即$C$满足定义11的题设条件，所以由定理1可知$C$既不是内涵也不是伪内涵。由于$C$是区间$\langle B,f((g(B))\rangle$内任意的属性集合，因此在区间$\langle B,f(g(B))\rangle$内，既不存在内涵也不存在伪内涵。

(3)设$N=f(g(B))$，因为字典序是一个线性序，所以$N\subseteq f(g(B))$，$f(g(B))\subseteq N$。由定义10可知，对于任意的属性集合$C\in\langle B,N\rangle$，都满足$B\subseteq C$。因为$C<N=f(g(B))$，所以由引理1可知$f(g(B))\not\subseteq C$，即C满足定义11的题设条件，所以由定理1可知C既不是内涵也不是伪内涵。由于C是区间$\langle B,N\rangle$内任意的属性集合，因此在区间$\langle B,N\rangle$内，既不存在内涵也不存在伪内涵。证毕。

定理2表明，对于主基中的任何一个蕴涵式，通过计算蕴涵式前件属性集合的仅大于属性集合的不相关性，以及非平凡仅大于属性集合的相关性，可以得到一个既不存在内涵也不存在伪内涵的属性集合区间。这为我们在以字典序遍历并判断属性集是否为伪内涵或内涵时，忽略前述这些属性集合区间的计算，提供了理论依据。

定理3 给定形式背景$K=(U,M,I)$与K上的主基$J(K)$，$T\subseteq M$，蕴涵式集合$J_{<T}(K)=\{C->D|C->D\in J(K)$且$C<T\}$。若属性集合$T$与$J_{<T}(K)$相关，则$T$与$J(K)$相关。

证明：由题干知$J_{<T}(K)\subseteq J(K)$，因为T与$J_{<T}(K)$相关，所以对于任意的$C->D\in J_{<T}(K)$，都满足$C\not\subseteq T$或$D\subseteq T$；又因为对于任意的$E->F\in J(K)-J_{<T}(K)$，都满足$T<E$，所以由引理1可知$E\not\subseteq T$，故T满足定义6的条件，即T与$J(K)-J_{<T}(K)$相关。又因为T与$J_{<T}(K)$相关，所以T与$J(K)$相关。证毕。

定理3表明，给定一个形式背景与主基，若某个属性集合与主基中小于其字典序的蕴涵式都相关，则该属性集合与主基相关。即某个属性集合与主基相关的必要条件是该属性集合与主基中小于其字典序的蕴涵式都相关。在属性探索算法中，在以字典序遍历属性集合时，只需要考虑此属性集合是否与当前的部分主基相关，不需要考虑其字典序之后的蕴涵式。定理3为我们在属性探索中判断属性集合与主基是否相关提供了理论依据。

3.2 属性探索算法的改进

本节在上述定义及定理的基础上，借鉴传统属性探索算法与专家问答的框架，设计了一种基于不相关属性集合的属性探索算法(AEUS)。该算法在以字典序遍历测试属性集合是否为内涵或伪内涵时，利用定理2直接跳过与主基不相关的属性集合。算法的描述如算法1所示。

算法1 AEUS算法

输入：属性集M，专家脑中关于属性依赖是否成立的判定(在运行中交互)

输出：主基$J_i(K)$，内涵集合$C_i(K)$

```
BEGIN
1. J0(K)←Ø;C0(K)←Ø;K0←Ø;U0←Ø;B0←Ø;
2. FLAG =FALSE
3. WHILE(Bi≠M)
4. 在Ki中计算 f(g(Bi))
5. 问专家Bi->f(g(Bi))-Bi在 K 中是否成立?
6. IF 不成立,专家给出反例ui
7. Ui+1←Ui∪ui
8. Bi+1←Bi
9. Ji+1(K)←Ji(K)
10. Ci+1(K)←Ci(K)
11. Ki+1←K(Ui+1,M,I)
12. ELSE
13. Ki+1←Ki
14.   IF(f(g(Bi))≠Bi)
15.   FLAG=TRUE
16. Ci+1(K)←Ci(K)
17. Ji+1(K)←Ji(K)∪(Bi->f(g(Bi))-Bi)
18.   ELSE
19. Ji+1(K)←Ji(K)
20.   Ci+1(K)←Ci(K)∪(Bi)
21.   END IF
22. B'←仅大于Bi的属性集合
23.   IF(B'与 J(K)相关) THEN
24. Bi+1←B'
25.   ELSE
26.   IF(FLAG) THEN
27. Bi+1←findNextB(Bi,Ji+1(K))
28.   ELSE
29. Bi+1←findNextB(B',Ji+1(K))
30.   END IF
31.   END IF
32. END IF
33. END WHILE
34. END
```

算法1初始时，形式背景为空，主基为空，内涵集为空。然后，不断以字典序产生要测试的属性集，询问专家以该属性集为前件的蕴涵式是否成立。如果不成立，则在形式背景中添加一个反例并重新计算；如果成立，则判断该属性集是内涵还是伪内涵。如果是伪内涵，则产生以该伪内涵为前件的蕴涵式并加入到主基中。如果不是伪内涵，根据概念格的值依赖与属性集合相关性理论，其必然是内涵，则将该属性集加入到内涵集。最后，按字典序产生下一个待测试的属性集，在此过程中将利用定理2的理论依据，直接跳过不需要测试的集合。算法具体步骤的分析如下。

AEUS算法的第4行、第5行在形式背景K_i中计算$f(g(B_i))$，并以此生成蕴涵关系式$B_i->f(g(B_i))-B_i$与领域专家进行交互问答。第6行、第7行，专家判断上述蕴涵式不成立，并根据自己拥有的知识体系，提供一个反例加入到形式背景K_i中，得到形式背景K_{i+1}。由于形式背景发生了改变，因此需要令$B_{i+1}\leftarrow B_i$，重新计算$f(g(B_i))$。第14行－第31行是专家判断蕴涵关系式成立时的处理过程。第14行判断

B_i是否为伪内涵，若 $f(g(B_i))\neq B_i$，则说明B_i是伪内涵，将蕴涵关系式加入主基。第 18 行、第 19 行说明B_i是内涵，将B_i加入内涵集合。至此算法对B_i的探索已经完成。第 22 行－第 31 行是算法根据B_i与当前的主基，计算下一个需要探索的属性集合B_{i+1}。首先计算出字典序中仅大于B_i的属性集合B'，由定义 9 可知区间$\langle B_i, B'\rangle=\emptyset$，如果$B'$与$J_i(K)$相关，由定理 3 可知$B'$与主基相关，则$B_{i+1}=B'$。若$B'$与当前主基不相关，则利用定理 2，跳过某个属性区间从而得到B_{i+1}。其中 findNextB 算法是利用定理 2 计算B_{i+1}的过程，具体描述如算法 2 所示。findNextB 算法第 1 行中，T 是B_i所有子集在形式背景中经过 f 运算与 g 运算后的并集，第 2 行计算出非平凡仅大于B_i的属性集合 N。该算法根据定理 2 跳过区间$\langle B_i, B_{i+1}\rangle$内的属性集合，从而计算出$B_{i+1}$。

算法 2 findNextB 算法描述

```
输入：属性集合Bi，主基 J(K)
输出：下一个属性集合Bi+1
1. T←∪f(g(Bi))
2. N←非平凡仅大于Bi的属性集合
3. WHILE(N 与 J(K)不相关)
4. N←非平凡仅大于 N 的属性集合
5. T←∪f(g(N)
6. 把 T 添加到 TArray 中
7. END WHILE
8. IF(TArray!=null) THEN
9. T←min(TArray)
10.   END IF
11. IF(T>N)THEN
12.   RETURN N ELSE
13.   RETURN T
14.   END IF
15. END
```

设形式背景的规模是 $m*n$，在 AEUS 算法中每个B_i都需要遍历一次形式背景来计算 $f(g(B_i))$，因此对于每个B_i来说，时间复杂度为 $O(m*n)$。findNextB 算法在计算B_{i+1}时每次最坏的情况需要计算 n 次 N，而每次判断 N 是否与主基相关需要遍历主基。因为主基的规模与 m 和 n 相关，但这个关系不是很明确，所以设主基的规模为 P，则每次计算 N 的时间复杂度是 $O(n*P)$，又因为计算 T 需要遍历主基，而且规模为 $m*n$ 的形式背景中有 n 个 N，所以 findNextB 算法最坏的时间复杂度为 $O(n*n*P*P)$，最好的情况下每次仅需计算一次 N，即时间复杂度为 $O(n*P*P)$。因此，AEUS 算法最坏的时间复杂度为 $O(m*n*n*n*P*P)$，最好的时间复杂度为 $O(m*n*n*P*P)$。虽然本文算法的时间复杂度与 P 相关，但是与传统属性探索算法的时间复杂度相比，本文算法的时间复杂度远小于传统属性探索算法。

4 属性探索算法的过程示例

本节通过一个示例来分别阐述文献[21]的属性探索算法（为方便阐述，本文将其记为 AERS 算法）与本文提出的 AEUS 算法的运行过程，重点对比上述两种算法在求解B_i的下一个属性集合B_{i+1}的过程中所需步骤的数量。给定形式背景 $K=(U,M,I)$，其中 $U=(1,2,3,4)$，$M=(a,b,c,d,e,f,g,h,i)$且 $a<b<c<d<e<f<g<h<i$。其中，专家脑中的知识背景 K 如表 1 所列。

表 1 形式背景 K

Table 1 Formal context K

	a	b	c	d	e	f	g	h	i
1	0	0	1	1	1	1	1	0	0
2	0	0	0	0	0	0	1	0	1
3	0	0	1	1	1	0	1	0	0
4	1	1	1	1	0	0	0	1	0

4.1 AERS 算法过程示例

初始状态$K_0=\emptyset$（见表 2），$C_0(K)=\emptyset$，$J_0(K)=\emptyset$，$B_0=\emptyset$。

表 2 形式背景K_0

Table 2 Formal context K_0

	a	b	c	d	e	f	g	h	i

（1）在形式背景K_0中，$f(g(\emptyset))=\{abcdefghi\}$，问专家 $\emptyset->f(g(\emptyset))-\emptyset$ 即 $\emptyset->abcdefghi$ 在 K 中是否成立。在 K 中，$g(\emptyset)=\{1,2,3,4\}$，$g(abcdefghi)=\emptyset$。因为$\{1,2,3,4\}\not\subseteq\emptyset$，所以 $\emptyset->abcdefghi$ 在 K 中不成立。从 K 中取出反驳蕴涵式 $\emptyset->abcdefghi$ 成立的反例对象 1，将其加入形式背景K_0，得到形式背景K_1（见表 3），$C_1(K)=C_0(K)$，$B_1=B_0$，$J_1(K)=J_0(K)$。

表 3 形式背景K_1

Table 3 Formal context K_1

	a	b	c	d	e	f	g	h	i
1	0	0	1	1	1	1	1	0	0

（2）在形式背景K_1中，$f(g(\emptyset))=\{cdefg\}$，问专家 $\emptyset->f(g(\emptyset))-\emptyset$ 即 $\emptyset->cdefg$ 在 K 中是否成立。在 K 中$g(\emptyset)=\{1,2,3,4\}$，$g(cdefg)=\{1\}$。因为$\{1,2,3,4\}\not\subseteq\{1\}$，所以 $\emptyset->cdefg$ 在 K 中不成立。从 K 中取出反驳蕴涵式 $\emptyset->cdefg$ 成立的反例对象 2，将其加入形式背景K_1，得到形式背景K_2（见表 4），$C_2(K)=C_1(K)$，$B_2=B_1$，$J_2(K)=J_1(K)$。

表 4 形式背景K_2

Table 4 Formal context K_2

	a	b	c	d	e	f	g	h	i
1	0	0	1	1	1	1	1	0	0
2	0	0	0	0	0	0	1	0	1

（3）在形式背景K_2中，$f(g(\emptyset))=\{g\}$，问专家 $\emptyset->f(g(\emptyset))-\emptyset$ 即 $\emptyset->g$ 在 K 中，是否成立。在 K 中，$g(\emptyset)=\{1,2,3,4\}$，$g(g)=\{1,2,3\}$。因为$\{1,2,3,4\}\not\subseteq\{1,2,3\}$，所以 $\emptyset->g$ 在 K 中不成立。从 K 中取出反驳蕴涵式 $\emptyset->g$ 成立的反例对象 4，将其加入形式背景K_2，得到形式背景K_3（见表 5），$C_3(K)=C_2(K)$，$B_3=B_2$，$J_3(K)=J_2(K)$。

表 5 形式背景K_3

Table 5 Formal context K_3

	a	b	c	d	e	f	g	h	i
1	0	0	1	1	1	1	1	0	0
2	0	0	0	0	0	0	1	0	1
4	1	1	1	1	0	0	0	1	0

(4)在形式背景K_3中，$f(g(\varnothing))=\{\varnothing\}$，问专家$\varnothing->f(g(\varnothing))-\varnothing$即$\varnothing->\varnothing$在$K$中是否成立。在$K$中，$g(\varnothing)=\{1,2,3,4\}$，$g(\varnothing)=\{1,2,3,4\}$。因为$\{1,2,3,4\}=\{1,2,3,4\}$，所以$\varnothing->\varnothing$在$K$中成立。又因为$f(g(\varnothing))=\varnothing$，所以$\varnothing$为内涵，$C_4(K)=C_3(K)\cup\varnothing$，$K_4=K_3$，$J_4(K)=J_3(K)$，计算$B_4$。

(5)$\varnothing$的下一个元素是$\{i\}$，$\{i\}$与$J_4(K)$相关，所以$B_4=i$。

(6)在形式背景K_4中，$f(g(i))=\{ig\}$，问专家$i->f(g(i))-i$即$i->g$在K中是否成立。在K中，$g(i)=\{1\}$，$g(g)=\{1,2,3\}$。因为$\{1\}\subseteq\{1,2,3\}$，所以$i->g$在K中成立。又因为$f(g(i))=\{ig\}\neq i$，所以i为伪内涵，$C_5(K)=C_4(K)$，$K_5=K_4$，$J_5(K)=J_4(K)\cup\{i->g\}$，计算$B_5$。

(7)i的下一个元素是$\{h\}$，$\{h\}$与$J_5(K)$相关，所以$B_5=h$。

(8)在形式背景K_5中，$f(g(h))=\{abcdh\}$，问专家$h->f(g(h))-h$即$h->abcd$在K中是否成立。在K中，$g(h)=\{4\}$，$g(abcd)=\{4\}$。因为$\{4\}=\{4\}$，所以$h->abcd$在K中成立。又因为$f(g(h))=\{abcdh\}\neq h$，所以h为伪内涵，$C_6(K)=C_5(K)$，$K_6=K_5$，$J_6(K)=J_5(K)\cup\{h->abcd\}$，计算$B_6$。

(9)h的下一个元素是$\{hi\}$，$\{hi\}$与$J_6(K)$不相关；hi的下一个元素是$\{g\}$，$\{g\}$与$J_6(K)$相关，所以$B_6=g$。

(10)在形式背景K_6中，$f(g(g))=\{g\}$，问专家$g->f(g(g))-g$即$g->\varnothing$在K中是否成立。在K中，$g(g)=\{1,2,3\}$，$g(\varnothing)=\{1,2,3,4\}$。因为$\{1,2,3\}\subseteq\{1,2,3,4\}$，所以$g->\varnothing$在$K$中成立。又因为$f(g(g))=\{g\}=g$，所以$g$为内涵，$C_7(K)=C_6(K)\cup\{g\}$，$K_7=K_6$，$J_7(K)=J_6(K)$，计算$B_7$。

(11)g的下一个元素是$\{gi\}$，$\{gi\}$与$J_7(K)$相关，所以$B_7=gi$。

(12)在形式背景K_7中，$f(g(gi))=\{gi\}$，问专家$gi->f(g(gi))-gi$即$gi->\varnothing$在K中是否成立。在K中，$g(gi)=\{2\}$，$g(\varnothing)=\{1,2,3,4\}$。因为$\{2\}\subseteq\{1,2,3,4\}$，所以$gi->\varnothing$在$K$中成立。又因为$f(g(gi))=\{gi\}=gi$，所以$gi$为内涵，$C_8(K)=C_7(K)\cup\{gi\}$，$K_8=K_7$，$J_8(K)=J_7(K)$，计算$B_8$。

(13)gi的下一个元素是$\{gh\}$，$\{gh\}$与$J_8(K)$不相关；gh的下一个元素是$\{ghi\}$，$\{ghi\}$与$J_8(K)$不相关；ghi的下一个元素是$\{f\}$，$\{f\}$与$J_8(K)$相关，所以$B_8=f$。

(14)在形式背景K_8中，$f(g(f))=\{cdefg\}$，问专家$f->f(g(f))-f$即$f->cdeg$在K中是否成立。在K中，$g(f)=\{1\}$，$g(cdeg)=\{1,3\}$。因为$\{1\}\subseteq\{1,3\}$，所以$f->cdeg$在K中成立。又因为$f(g(f))=\{cdeg\}\neq f$，所以f为伪内涵，$C_9(K)=C_8(K)$，$K_9=K_8$，$J_9(K)=J_8(K)\cup\{f->cdeg\}$，计算$B_9$。

(15)f的下一个元素是$\{fi\}$，$\{fi\}$与$J_9(K)$不相关；fi的下一个元素是$\{fh\}$，$\{fh\}$与$J_9(K)$不相关；fh的下一个元素是$\{fhi\}$，$\{fhi\}$与$J_9(K)$不相关；fhi的下一个元素是$\{fg\}$，$\{fg\}$与$J_9(K)$不相关；fg的下一个元素是$\{fgi\}$，$\{fgi\}$与$J_9(K)$不相关；fgi的下一个元素是$\{fgh\}$，$\{fgh\}$与$J_9(K)$不相关；fgh的下一个元素是$\{fghi\}$，$\{fghi\}$与$J_9(K)$不相关；$fghi$的下一个元素是$\{e\}$，$\{e\}$与$J_9(K)$相关，所以$B_9=e$。即区间$\langle f,e\rangle$与$J_9(K)$不相关，$\{e\}$与$J_9(K)$相关。

(16)在形式背景K_9中，$f(g(e))=\{cdefg\}$，问专家$e->f(g(e))-e$即$e->cdfg$在K中是否成立。在K中，$g(e)=\{1,3\}$，$g(cdfg)=\{1\}$。因为$\{1,3\}\not\subseteq\{1\}$，所以$e->cdfg$在$K$中不成立。从$K$中取出反驳蕴涵式$e->cdfg$成立的反例对象3，将其加入形式背景$K_9$，得到形式背景$K_{10}$，如表6所列。

表6 形式背景K_{10}

Table 6 Formal context K_{10}

	a	b	c	d	e	f	g	h	i
1	0	0	1	1	1	1	1	0	0
2	0	0	0	0	0	0	1	0	1
4	1	1	1	1	0	0	0	1	0
3	0	0	1	1	1	0	1	0	0

(17)在形式背景K_{10}中，$f(g(e))=\{cdeg\}$，问专家$e->f(g(e))-e$即$e->cdg$在K中是否成立。在K中，$g(e)=\{1,3\}$，$g(cdeg)=\{1,3\}$。因为$\{1,3\}=\{1,3\}$，所以$e->cdeg$在K中成立。又因为$f(g(e))=\{cdeg\}\neq e$，所以e为伪内涵，$C_{10}(K)=C_9(K)$，$K_{10}=K_9$，$J_{10}(K)=J_9(K)\cup\{e->cd\}$。

(18)AERS算法计算B_{i+1}的过程如表7所列。

表7 AERS计算B_{i+1}的过程

Table 7 Calculate the B_{i+1} process of AERS

i	B_{i+1}
9	区间$\langle e,d\rangle$内属性集合都与$J_{10}(K)$不相关，$\{d\}$与$J_{10}(K)$相关，所以$B_{10}=d$
10	区间$\langle d,c\rangle$内属性集合都与$J_{11}(K)$不相关，$\{c\}$与$J_{11}(K)$相关，所以$B_{11}=c$
11	区间$\langle c,cd\rangle$内属性集合都与$J_{12}(K)$不相关，$\{cd\}$与$J_{12}(K)$相关，所以$B_{12}=cd$
12	区间$\langle cd,cdg\rangle$内属性集合都与$J_{13}(K)$不相关，$\{cdg\}$与$J_{13}(K)$相关，所以$B_{13}=cdg$
13	区间$\langle cdg,cdeg\rangle$内属性集合都与$J_{14}(K)$不相关，$\{cdeg\}$与$J_{14}(K)$相关，所以$B_{14}=cdeg$
14	$\{cdegi\}$与$J_{15}(K)$相关，所以$B_{15}=cdegi$
15	区间$\langle cdegi,cdefg\rangle$内属性集合都与$J_{16}(K)$不相关，$\{cdefg\}$与$J_{16}(K)$相关，所以$B_{16}=cdefg$
16	区间$\langle cdefg,b\rangle$内属性集合都与$J_{17}(K)$不相关，$\{b\}$与$J_{17}(K)$相关，所以$B_{17}=b$
17	区间$\langle b,a\rangle$内属性集合都与$J_{18}(K)$不相关，$\{a\}$与$J_{18}(K)$相关，所以$B_{18}=a$
18	区间$\langle a,abcdh\rangle$内属性集合都与$J_{19}(K)$不相关，$\{abcdh\}$与$J_{19}(K)$相关，所以$B_{19}=abcdh$
19	区间$\langle abcdh,abcdegh\rangle$内属性集合都与$J_{20}(K)$不相关，$\{abcdegh\}$与$J_{20}(K)$相关，所以$B_{20}=abcdegh$
20	区间$\langle abcdegh,abcdefghi\rangle$内属性集合都与$J_{21}(K)$不相关，$\{abcdefghi\}$与$J_{21}(K)$相关，所以$B_{21}=abcdefghi$
21	结束

由上述过程可以看出，AERS算法中区间内属性集合的数量非常多。因为该算法需要遍历较多的属性集合，所以该算法的时间复杂度高。

4.2 AEUS算法过程示例

由于AEUS算法与AERS算法的计算情形从步骤(1)—步骤(4)均相同，因此省略该步骤，从步骤(5)开始讨论。

(5)$\varnothing$的下一个元素$B'=\{i\}$，B'与$J_4(K)$相关，所以$B_4=i$。

(6)在形式背景K_4中$f(g(i))=\{ig\}$，问专家$i->f(g$

$(i))-i$ 即 $i->g$ 在 K 中是否成立。在 K 中 $g(i)=\{1\}$，$g(g)=\{1,2,3\}$。因为 $\{1\}\subseteq\{1,2,3\}$，所以 $i->g$ 在 K 中成立。又因为 $f(g(i))=\{ig\}\neq i$，所以 i 为伪内涵，$C_5(K)=C_4(K)$，$K_5=K_4$，$J_5(K)=J_4(K)\cup\{i->g\}$，计算 B_5。

(7) i 的下一个元素 $B'=\{h\}$，B' 与 $J_5(K)$ 相关，所以 $B_5=h$。

(8)在形式背景 K_5 中，$f(g(h))=\{abcdh\}$，问专家 $h->f(g(h))-h$ 即 $h->abcd$ 在 K 中是否成立。在 K 中，$g(h)=\{4\}$，$g(abcd)=\{4\}$。因为 $\{4\}=\{4\}$，所以 $h->abcd$ 在 K 中成立。又因为 $f(g(h))=\{abcdh\}\neq h$，所以 h 为伪内涵，$C_6(K)=C_5(K)$，$K_6=K_5$，$J_6(K)=J_5(K)\cup\{h->abcd\}$，计算 B_6。

(9) h 字典序下一个元素 $B'=\{hi\}$，B' 与 $J_6(K)$ 不相关，计算 $T=abcdh$，$N=g$，因为 $T>N$，所以 $B_6=g$。

(10)在形式背景 K_6 中，$f(g(g))=\{g\}$，问专家 $g->f(g(g))-g$ 即 $g->\varnothing$ 在 K 中是否成立。在 K 中，$g(g)=\{1,2,3\}$，$g(\varnothing)=\{1,2,3,4\}$。因为 $\{1,2,3\}\subseteq\{1,2,3,4\}$，所以 $g->\varnothing$ 在 K 中成立。又因为 $f(g(g))=\{g\}=g$，所以 g 为内涵，$C_7(K)=C_6(K)\cup\{g\}$，$K_7=K_6$，$J_7(K)=J_6(K)$，计算 B_7。

(11) g 的下一个元素 $B'=\{gi\}$，B' 与 $J_7(K)$ 相关，所以 $B_7=gi$。

(12)在形式背景 K_7 中，$f(g(gi))=\{gi\}$，问专家 $gi->f(g(gi))-gi$ 即 $gi->\varnothing$ 在 K 中是否成立。在 K 中，$g(gi)=\{2\}$，$g(\varnothing)=\{1,2,3,4\}$。因为 $\{2\}\subseteq\{1,2,3,4\}$，所以 $gi->\varnothing$ 在 K 中成立。又因为 $f(g(gi))=\{gi\}=gi$，所以 gi 为内涵，$C_8(K)=C_7(K)\cup\{gi\}$，$K_8=K_7$，$J_8(K)=J_7(K)$，计算 B_8。

(13) gi 的下一个元素 $B'=\{gh\}$，B' 与 $J_8(K)$ 不相关，计算 $T=abcdgh$，$N=f$。因为 $N<T$，所以 $B_8=N=f$。

(14)在形式背景 K_8 中，$f(g(f))=\{cdefg\}$，问专家 $f->f(g(f))-f$ 即 $f->cdeg$ 在 K 中是否成立。在 K 中，$g(f)=\{1\}$，$g(cdeg)=\{1,3\}$。因为 $\{1\}\subseteq\{1,3\}$，所以 $f->cdeg$ 在 K 中成立。又因为 $f(g(f))=\{cdeg\}\neq f$，所以 f 为伪内涵，$C_9(K)=C_8(K)$，$K_9=K_8$，$J_9(K)=J_8(K)\cup\{f->cdeg\}$，计算 B_9。

(15) f 的下一个元素 $B'=\{fi\}$，$\{fi\}$ 与 $J_9(K)$ 不相关，计算 $T=cdefg$，$N=e$，因为 $T>e$，所以 $B_9=e$。

(16)在形式背景 K_9 中，$f(g(e))=\{cdefg\}$，问专家 $e->f(g(e))-e$ 即 $e->cdfg$ 在 K 中是否成立。在 K 中，$g(e)=\{1,3\}$，$g(cdfg)=\{1\}$。因为 $\{1,3\}\not\subseteq\{1\}$，所以 $e->cdfg$ 在 K 中不成立。从 K 中取出反驳蕴涵式 $e->cdfg$ 成立的反例 3，将其加入形式背景 K_9，得到形式背景 K_{10}，如表 8 所列。

表 8 形式背景 K_{10}

Table 8 Formal context K_{10}

	a	b	c	d	e	f	g	h	i
1	0	0	1	1	1	1	1	0	0
2	0	0	0	0	0	0	1	0	1
4	1	1	1	1	0	0	0	1	0
3	0	0	1	1	1	0	1	0	0

(17)在形式背景 K_{10} 中，$f(g(e))=\{cdeg\}$，问专家 $e->f(g(e))-e$ 即 $e->cdg$ 在 K 中是否成立。在 K 中，$g(e)=\{1,3\}$，$g(cdeg)=\{1,3\}$。因为 $\{1,3\}=\{1,3\}$，所以 $e->cdeg$ 在 K 中成立。又因为 $f(g(e))=\{cdeg\}\neq e$，所以 e 为伪内涵，$C_{10}(K)=C_9(K)$，$K_{10}=K_9$，$J_{10}(K)=J_9(K)\cup\{e->cdg\}$。

(18)AEUS 算法计算 B_{i+1} 的过程如表 9 所列。

表 9 AEUS 计算 B_{i+1} 的过程

Table 9 Calculate the B_{i+1} process of AEUS

i	B_{i+1}
9	$B'=ei$，与 $J_{10}(K)$ 不相关，计算 $T=cdeg$，$N=d$，$T>d$，所以 $B_{10}=d$
10	$B'=di$，与 $J_{11}(K)$ 不相关，计算 $T=cd$，$N=c$，$T>c$，所以 $B_{11}=c$
11	$B'=ci$，与 $J_{12}(K)$ 不相关，计算 $T=cd$，$N=b$，$b>T$，所以 $B_{12}=cd$
12	$B'=cdi$，与 $J_{13}(K)$ 不相关，计算 $TArray=\{cdgi,abcdh\}$，$N=cdg$，$TArray$ 最小的属性集合 $cdgi>cdg$，所以 $B_{13}=cdg$
13	$B'=cdgi$，与 $J_{14}(K)$ 不相关，计算 $T=cdeg$，$N=b$，$T<b$，所以 $B_{14}=cdeg$
14	$B'=cdegi$，与 $J_{15}(K)$ 相关，所以 $B_{15}=cdegi$
15	$B'=cdgh$，与 $J_{16}(K)$ 不相关，计算 $TArray=\{abcdeghi\}$，$N=cdefg$，$TArray$ 最小的属性集合 $cdefg<abcdeghi$，所以 $B_{16}=cdefg$
16	$B'=cdefgi$，与 $J_{17}(K)$ 不相关，计算 $TArray=\{abcdefgh,abcdefghi\}$，$N=b$，$TArray$ 最小的属性集合 $abcdefgh>b$，所以 $B_{17}=b$
17	$B'=bi$，与 $J_{18}(K)$ 不相关，计算 $T=abcdh$，$N=a$，$abcdh>a$，所以 $B_{18}=a$
18	$B'=ai$，与 $J_{19}(K)$ 不相关，计算 $T=abcdh$，所以 $B_{19}=abcdh$
19	$B'=abcdhi$，与 $J_{20}(K)$ 不相关，计算 $TArray=\{abcdegh,abcdefghi\}$，$TArray$ 最小的属性集合 $abcdegh$，所以 $B_{20}=abcdegh$
20	$B'=abcdeghi$，与 $J_{21}(K)$ 不相关，计算 $TArray=\{abcdefghi\}$，$TArray$ 最小的属性集合 $abcdefghi$，所以 $B_{21}=abcdefghi$
21	结束

从以上两个过程可以看出，AERS 算法在计算 B_{i+1} 时需要遍历属性集合 M 的所有子集。当 M 较大时，会导致算法的搜索空间十分庞大。AEUS 算法可以避免逐个遍历属性集合的子集，从而减小算法搜索空间，达到缩短算法运行时间的目的。

5 实验与分析

5.1 实验设计

为验证本文改进算法的性能，使用 JAVA 语言 MATH 库中的 random 函数仿真生成一组形式背景作为测试数据。将本文的改进算法(AEUS)与传统属性探索算法[1](下文记为 TAE)、AERS 算法[21]进行对比实验。实验分为 3 个方面：1)改变实验条件，观察给定形式背景的蕴涵关系式数量；2)改变实验条件，对上述 3 种算法的耗时情况进行对比；3)改变实验条件，观察 AEUS 算法跳过属性集合的个数与总集合数目的比值。

在实验中，以算法遍历形式背景的方式代替专家回答问题。算法以随机生成的形式背景为专家所拥有的知识，在判断蕴涵关系式是否成立时，遍历整个形式背景，如果形式背景中所有对象满足此条蕴涵式的蕴涵关系，则认为这条蕴涵式成立；否则认为该条蕴涵关系式不成立，在形式背景中取出一个对象作为反例提供给算法。上述 3 种算法均以此种方式代替专家回答，因此这不会影响实验对比结果。测试平台的硬件为 3.4 GHz 的 CPU 和 16 GB 内存，操作系统为 Windows x10，测试平台软件为 JDK9，Eclipse。

第 1 组实验设置形式背景具有相同的对象数目 50，属性

数目从 0 以间隔 5 变化到 30。测试目的是固定对象数目，改变属性个数，观察蕴涵式数量的变化。测试结果如图 1 所示。

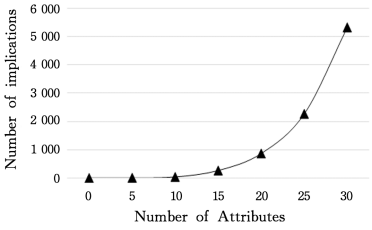

图 1 蕴涵式个数(对象个数为 50)

Fig. 1 Number of implications(number of objects is 50)

第 2 组实验设置形式背景具有相同的属性数目 15，对象数目从 0 以间隔 50 变化到 300。测试目的是固定属性数目，改变对象数目，观察蕴涵式数量的变化。测试结果如图 2 所示。

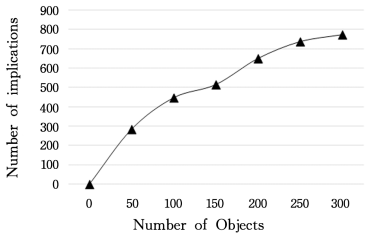

图 2 蕴涵式个数(属性个数为 15)

Fig. 2 Number of implications(number of attributes is 15)

第 3 组实验设置形式背景具有相同的对象数目 50，属性数目从 0 以间隔 5 变化到 30。测试目的是固定对象数目，改变属性个数，观察上述 3 种算法的耗时情况。测试结果如图 3 所示。

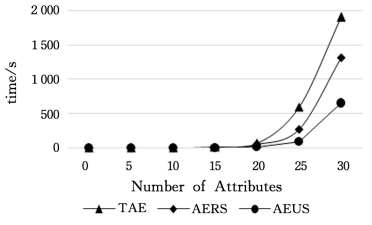

图 3 效率对比(对象个数为 50)

Fig. 3 Efficiency comparison(number of objects is 50)

第 4 组实验设置形式背景具有相同的属性数目 15，对象数目从 0 以间隔 50 变化到 300。测试目的是固定属性个数，改变对象数目，观察上述 3 种算法的耗时情况。测试结果如图 4 所示。

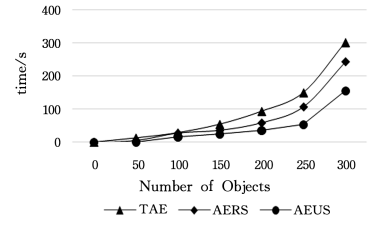

图 4 效率对比(属性个数为 15)

Fig. 4 Efficiency comparison(number of attributes is 15)

第 5 组实验设置形式背景属性与对象具有相等的数目，数目从 0 以间隔 5 变化到 30。测试目的是改变形式背景规模，观察 AEUS 算法的优化效率。测试结果如图 5 所示。

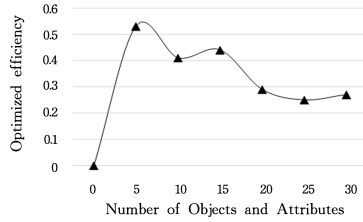

图 5 优化效率

Fig. 5 Optimized efficiency

5.2 实验分析

从第 1 组、第 2 组实验可以看出，给定形式背景，蕴涵关系式的个数随着属性、对象的个数增加而增加。同时我们对比 TAE，AERS 与 AEUS 得到的主基发现，这 3 种算法得到的结果是一致的，这两组实验也从侧面表明了 AEUS 算法的正确性。

第 3 组、第 4 组实验表明，不管是固定对象个数、改变属性个数，还是固定属性个数、改变对象个数，本文提出的改进算法的耗时都低于对比的两种算法，并且属性的数目越多，本文算法节约的时间越长。其中，在实验 3 中，当属性数目为 30 时，AEUS 算法的耗时仅是 AERS 算法的 50%。在实验 4 中，当对象数目为 300 时，AEUS 算法的耗时仅是 AERS 算法的 64%。

第 5 组实验表明，AEUS 算法可以有效地减小属性探索算法计算下一个需要探索属性集合时的搜索空间，减少的属性集合数目约为 25%～50%。

实验结果说明，本文提出的 AEUS 算法可以有效降低属性探索算法的时间复杂度。

结束语 针对现有的属性探索算法需要逐个遍历字典序集合，造成时间复杂度高的问题，我们希望算法自动跳过某些不可能成为内涵或伪内涵的属性集合，从而降低算法的时间复杂度。研究发现，存在一类与主基不相关的集合，这些集合包含主基中某个蕴涵式的前件，但不包含这个蕴涵式的后件，而且主基与内涵集合都不包含这类属性集合。本文将这种内在逻辑关系定义为不相关关系，进一步归纳总结出 3 个定理，并对提出的定理进行了严谨的数学论证。最后根据定理给出了一种改进的属性探索算法(AEUS)，该算法借助上述定理，在计算属性探索下一个需要探索的属性集合时，自动跳过与主基不相关的属性集合，改进了属性探索算法最为耗时的一步，减小了属性探索算法的搜索空间。因为属性探索算法可以半自动化辅助学习系统知识，接下来我们将属性探索算法进一步开发为更易操作的工具；同时将本文提出的 AEUS 算法进行并行处理，从而进一步改善属性探索算法的高耗时问题。

参 考 文 献

[1] GANTER B, WILLE R. Formal concept analysis: mathematical foundations[M]. Springer Science & Business Media, 2012.

[2] LI J H, WEI L, ZHANG Z, et al. Concept lattice theory and

method and their research prospect [J]. Pattern Recognition and Artificial Intelligence, 2020, 33(7): 619-642.
[3] JABBARI S, STOFFEL K. A Methodology for Extracting Knowledge about Controlled Vocabularies from Textual Data using FCA-Based Ontology Engineering[C]// 2018 IEEE International Conference on Bioinformatics and Biomedicine(BIBM). Madrid Spain, 2018: 1657-1661.
[4] MAHANI A, BABA-ALI A R. A new rule-based knowledge extraction approach for imbalanced datasets[J]. Knowledge and Information Systems, 2019, 61: 1303-1329.
[5] WEI L, LIU L, QI J J, et al. Rules acquisition of formal decision contexts based on three-way concept lattices[J]. Information Sciences, 2020, 516: 529-544.
[6] MI Y L, LIU W Q, SHI Y, et al. Semi-supervised Concept Learning by Concept-cognitive Learning and Concept Space[J]. IEEE Transactions on Knowledge and Data Engineering, 2020 (99): 1-1.
[7] ZHI H L, LI Y N. Knowledge representation base on concept cluster[J]. journal of Northwest University(Nature Science Edition), 2020, 50(4): 529-536.
[8] LI J H, HE J J, WU W Z. Optimization of class-attribute block in multi-granularity formal concept analysis[J]. Journal of Shandong University(Natural Science), 2020, 55(5): 1-12.
[9] GANTER B, OBIEDKOV S, RUDOLPH S, et al. Conceptual exploration[M]. Heidelberg: Springer, 2016.
[10] BORCHMANN D. A general form of attribute exploration[J]. arXiv: 1202. 4824, 2012.
[11] BORCHMANN D. Exploring faulty data[C]// International Conference on Formal Concept Analysis. Springer, Cham, 2015: 219-235.
[12] GLODEANU C V. Attribute Exploration with Fuzzy Attributes and Background Knowledge[C]// CLA. 2013: 69-80.
[13] OBIEDKOV S, KOURIE D G, ELOFF J H P. Building access control models with attribute exploration[J]. Computers & Security, 2009, 28(1/2): 2-7.
[14] ROBERT J, SEBASTIAN R. Attribute Exploration on the Web [C]// the 11th International Conference on Formal Concept Analysis. 2013: 19-34.
[15] OBIEDKOV S, ROMASHKIN N. Collaborative conceptual exploration as a tool for crowdsourcing domain ontologies[C]// Proceedings of Russian and South African Workshop on Knowledge Discovery Techniques Based on Formal Concept Analysis, CEUR Workshop Proceedings. 2015, 1552: 58-70.
[16] HANIKA T, ZUMBRÄGEL J. Towards collaborative conceptual exploration[C]// International Conference on Conceptual Structures. Springer, Cham, 2018: 120-134.
[17] CODOCEDO V, BAIXERIES J, KAYTOUE M, et al. Sampling Representation Contexts with Attribute Exploration[C]// International Conference on Formal Concept Analysis. Springer, Cham, 2019: 307-314.
[18] RYSSEL U, DISTEL F, BORCHMANN D. Fast algorithms for implication bases and attribute exploration using proper premises[J]. Annals of Mathematics and Artificial Intelligence, 2014, 70(1/2): 25-53.
[19] OLLBOLD J, KÖHLING R, BORCHMANN D. Attribute exploration with proper premises and incomplete knowledge applied to the free radical theory of ageing[C]// International Conference on Formal Concept Analysis. Springer, Cham, 2014: 268-283.
[20] KRIEGEL F. Parallel attribute exploration [C]// International Conference on Conceptual Structures. Springer, Cham, 2016: 91-106.
[21] ZHAO X X, QIN P, WANG J. Research on attribute exploration algorithm[J]. Computer Science and Exploration, 2009, 3(5): 509-518.

SHEN Xia-jiong, born in 1963, Ph. D, professor, Ph.D supervisor, is a member of China Computer Fededration. His main research interests include data analysis, software engineering, distributed/parallel computing, formal concept analysis, access control and so on.

ZHANG Lei, born in 1981, Ph.D, associate professor. His main research interests include machine learning, big data, information security, access control, formal concept analysis and so on.