

## 基于修正 BPSO 的通用模式指标上界估算方法

王 菊<sup>1</sup>, 刘付显<sup>1</sup>, 靳春杰<sup>2</sup>

(1. 空军工程大学 防空反导学院, 西安 710051; 2. 93527 部队, 河北 张家口 075000)

**摘 要:** 针对约束频繁模式挖掘中模式指标的界值估算问题, 提出一种基于修正二进制粒子群优化(BPSO)算法的通用模式指标上界估算方法。根据带有权值的不确定型事务数据库的特点, 建立通用的模式指标上界估算框架, 并提出在该框架下基于修正 BPSO 的模式指标上界值求解方法。对比 UHUI-Apriori 算法分别结合事务加权效用值、本文方法估算所得上界值和实际上界值后的候选项集数量、运行时间和内存占用情况, 结果表明, 该方法可以较快计算模式效用的上界值, 且能够节省运行时间和内存空间。

**关键词:** 不确定型数据库; 模式指标; 界值估算; 粒子群优化算法; 约束频繁模式挖掘

**中文引用格式:** 王 菊, 刘付显, 靳春杰. 基于修正 BPSO 的通用模式指标上界估算方法[J]. 计算机工程, 2018, 44(10): 168-174.

**英文引用格式:** WANG Ju, LIU Fuxian, JIN Chunjie. Estimation method for upper bound of general pattern index based on modified BPSO[J]. Computer Engineering, 2018, 44(10): 168-174.

## Estimation Method for Upper Bound of General Pattern Index Based on Modified BPSO

WANG Ju<sup>1</sup>, LIU Fuxian<sup>1</sup>, JIN Chunjie<sup>2</sup>

(1. Air and Missile Defense College, Air Force Engineering University, Xi'an 710051, China;

2. 93527 Troop, Zhangjiakou, Hebei 075000, China)

**[Abstract]** A general pattern index upper bound estimation method based on modified Binary Particle Swarm Optimization (BPSO) algorithm is proposed to estimate the boundary value estimation problem in constrained frequent pattern mining. According to the characteristics of the uncertain transaction database with weights, a general pattern index upper bound estimation framework is established, and the method of solving the upper bound value of the pattern index based on the modified BPSO is proposed. The number of candidate itemses, runtime and memory usage of the UHUI-Apriori with Transaction Weighted Utilization(TWU), the proposed method estimates upper bound and the actual upper bound are compared, and the results show that the proposed method can quickly calculate the upper bound value of the mode utility and save the running time and memory space.

**[Key words]** uncertain database; pattern index; bound value estimation; Particle Swarm Optimization(PSO) algorithm; constrained frequent pattern mining

**DOI:**10.19678/j.issn.1000-3428.0047966

### 0 概述

作为频繁模式挖掘的拓展, 约束频繁模式挖掘近年来引起了广泛关注<sup>[1-2]</sup>。目前, 对其进行的研究主要集中在如何将不同的模式指标, 尤其是针对难约束的指标<sup>[3-4]</sup>, 融入到算法的执行过程中, 以减少搜索空间。其中, 一个通用处理办法就是估计给定项集  $X$  的所有超集关于模式指标  $PN$  的一个上(下)界。如果这个界值小于(大于)一个给定的最小(最大)阈值, 那么  $X$  的所有超集就要从全集中被剪

除<sup>[5-6]</sup>。因此, 模式指标的界值估算是约束频繁模式挖掘中一个非常重要的问题。

当前, 虽然在实际应用的驱动下, 一些特定模式指标及其上界估算方法已经被提出, 如模式效用<sup>[7-8]</sup>、模式占有度<sup>[9]</sup>和有权置信度<sup>[10]</sup>等。但是, 这些研究的局限性在于它们不具有普遍性, 即每当一个新的难约束指标被提出时, 都需要花费很大代价去找到相应有效的上(下)界。因此, 文献[11]提出一种基于带有权值的确定型事务数据库的通用界值估算方法。但是, 该方法没有指明在带有权值的不

确定型事务数据库上如何进行运算。

针对上述问题,本文提出一种基于修正二进制粒子群优化(Binary Particle Swarm Optimization,BPSO)算法的通用模式指标上界值估算方法,该方法在传统界值估算方法的基础上增加了处理不确定性问题的功能。通过采用不考虑项标记而只考虑支持数据库中出现项的权值和存在概率的策略,建立面向带有权值的不确定型事务数据库的通用模式指标上界估算框架。在该框架的基础上,为快速搜索模式指标的上界值,利用 BPSO 算法机理简单、收敛速度快、全局优化能力强等方面的优势,提出基于修正 BPSO 的模式指标上界值求解方法。最后通过实验验证该方法的估算性能。

### 1 离散粒子群优化算法

1995 年,Kennedy 博士和 Eberhart 博士受到自然界中鸟群运动模型启发,提出一种新的基于种群的搜索算法——粒子群优化算法。该算法的设计思想来源于社会认知理论,可以将其求解过程划分为评价、比较和模仿 3 个阶段。相比于其他智能优化算法,该算法具有机理简单、收敛速度快、全局优化能力强等优势。

PSO 算法最初应用于连续空间的优化,相关研究也只集中在连续函数方面。但是在实际应用中,很多工程问题都是离散的,因此,亟需将基本 PSO 算法向二进制空间进行扩展。针对该问题,Kennedy 博士和 Eberhart 博士在 1997 年又进一步提出一种针对 0-1 规划的 BPSO 算法<sup>[12]</sup>。在 BPSO 算法中,一个二进制的空间就表示一个超立方体空间,每个粒子用一个二进制变量来表示,可以通过二进制变量的某些位在 {0,1} 之间的翻转来实现粒子在这个超立方体空间中的移动。该过程数学描述如下:

设粒子群的种群规模为  $M$ ,决策空间为  $n$  维,任意一个粒子  $i$  在第  $k$  代时的速度可以表示为  $V_i^k = (v_{i1}^k, v_{i2}^k, \dots, v_{id}^k, \dots, v_{in}^k), i = 1, 2, \dots, M, d = 1, 2, \dots, n$ ,位置可以表示为  $X_i^k = (x_{i1}^k, x_{i2}^k, \dots, x_{id}^k, \dots, x_{in}^k)$ ,且  $x_{id}^k \in \{0,1\}$ 。粒子  $i$  在第  $k+1$  代的第  $d$  维子空间中的飞行速度和位置根据式(1)、式(2)进行调整:

$$v_{id}^{k+1} = \begin{cases} \omega v_{id}^k + c_1 r_1 (p_{id} - x_{id}^k) + c_2 r_2 (p_{gd} - x_{id}^k), \\ -v_{\max} \leq v_{id}^{k+1} \leq v_{\max} \\ v_{\max}, v_{id}^{k+1} > v_{\max} \\ -v_{\max}, v_{id}^{k+1} < -v_{\max} \end{cases} \quad (1)$$

$$x_{id}^{k+1} = \begin{cases} 1, rand() < sig(v_{id}^{k+1}) \\ 0, else \end{cases} \quad (2)$$

其中, $\omega$  是惯性权重, $c_1, c_2$  是 2 个学习因子, $r_1, r_2$  和  $rand()$  都是  $[0,1]$  区间内的均匀随机数, $p_{id}$  是粒子  $i$  在历代中最优值的第  $d$  维, $p_{gd}$  是种群中全局最优个

体的第  $d$  维,函数  $sig(v_{id}^{k+1})$  是一个转换限制函数,能够使  $v_{id}^{k+1}$  值越大,粒子的位置  $x_{id}^{k+1}$  选 1 的概率就越大, $v_{id}^{k+1}$  值越小,粒子的位置  $x_{id}^{k+1}$  选 0 的概率就越大,其表达式为:

$$sig(v_{id}^{k+1}) = \frac{1}{1 + e^{-v_{id}^{k+1}}} \quad (3)$$

为防止  $sig(v_{id}^{k+1})$  函数饱和,通常设  $|v_{\max}| = 4$ ,以保证  $0.018 \leq sig(v_{id}^{k+1}) \leq 0.982$ 。

## 2 通用模式指标上界估算框架

### 2.1 模式指标定义

目前,关于模式指标上界的估算方法大多基于确定型数据库,为增加所提模型的通用性,本文的研究内容主要在带有权值的不确定型事务数据库  $D = \{T_1, T_2, \dots, T_m\}$  上展开。在该类数据库中,有限集合  $I = \{i_1, i_2, \dots, i_L\}$  表示  $D$  中包含的所有项集合,一个事务  $T_q = \{i_1, i_2, \dots, i_l\} (1 \leq l \leq L)$  包含  $I$  中有限数目的项,任意一个项  $i_p$  都对应于一个权值  $u(i_p, T_q)$  和一个存在概率  $p(i_p, T_q)$ ,它们分别表示  $i_p$  在事务  $T_q$  中的权值和  $i_p$  存在于事务  $T_q$  中的概率。表 1 所示为  $D$  的一个示例。

表 1 带有权值的不确定型事务数据库示例

事务编号	事务数据库
$T_1$	(A,3,0.81)(C,5,0.79)(D,8,0.93)(F,16,0.73)(G,12,0.85)
$T_2$	(B,24,0.52)(C,15,0.89)(D,24,0.86)
$T_3$	(A,6,0.93)(E,12,0.76)
$T_4$	(A,18,0.88)(C,11,0.92)(G,9,0.72)
$T_5$	(C,10,0.87)(F,9,0.73)
$T_6$	(A,25,0.83)(C,17,0.91)(D,24,0.86)(E,4,0.76)
$T_7$	(A,6,0.81)(D,8,0.92)(F,3,0.78)
$T_8$	(B,24,0.69)(C,30,0.83)(E,12,0.88)(F,6,0.63)

在实际应用中,在数据库  $D$  上以高概率出现且具有高权值的项集对决策者具有重要意义。为发现这类项集,本文定义了模式指标  $PN$  来综合评估项集的权值和存在概率,其主要包含 3 个输入:项集的支持数据库,每个项在其出现事务上的权值以及存在概率。模式指标  $PN$  的数学表达式为:

$$PN(X, D_X, H) : 2^L \times 2^m \times (2^L \times 2^m \rightarrow R_0^+) \rightarrow R_0^+ \quad (4)$$

其中, $X$  表示数据库  $D$  中的任意一个项集, $D_X$  表示项集  $X$  的支持数据库, $H = U \times P$  表示  $D_X$  中项的权值与其相应存在概率的乘积, $U$  表示  $D_X$  中项到权值的映射, $P$  表示  $D_X$  中项到存在概率的映射, $H$  矩阵中的值被称为概率权值。

通常情况下,可以利用模式指标的一些性质来估算其上界,如单调性、简洁性、可转换单调性、松散反单调性、灵活性、有界性等。但是在实际应用过程中,仍然存在一类不满足所有这些性质的约束(这类

约束被统称为难约束)。为降低计算复杂度、减小搜索空间,对难约束的一个通用处理办法就是估计给定项集  $X$  的所有父项集关于给定模式指标  $PN$  的一个上界。如果这个界值小于一个给定的最小阈值,则  $X$  的所有超集就要从全集中被剪除。当前针对某个或某几个特定的难约束,相关学者分别提出了其相应的上界,但是这些研究的局限性在于当一个新的难约束指标被提出时,都需要花费很大的代价去找到相应的有效上界。

因此,本文通过分析带有权值的不确定型事务数据库的特点和之前关于特定难约束指标上界估计方法的相同点,构建一种通用的模式指标估算框架。

## 2.2 基本框架

由于本文涉及的符号较多,本节首先就这些符号进行说明,如表 2 所示。

表 2 符号说明

符号名称	含义
$LT$	由 $D$ 中每个事务包含项的个数所构成的向量
$X'$	$X$ 的任意一个超集
$\alpha$	可能包含项集 $X'$ 的事务个数
$\beta$	$X'$ 相对于 $X$ 的扩展长度,满足 $\beta =  X'  -  X $
$XS$	由项集 $X$ 的支持事务编号所构成的向量
$ XS $	$XS$ 中所包含项的个数
$XS_\beta$	由相对于 $X$ 扩展长度大于 $\beta$ 的项集支持事务编号所构成的向量
$ XS_\beta $	$XS_\beta$ 中所包含项的个数
$ X $	$X$ 中所包含项的个数
$Pu$	$X$ 的前缀项集概率权值矩阵,大小为 $m \times  X $
$LS$	由去除 $X$ 后其支持事务中包含项的个数所构成的向量
$ LS $	$LS$ 中所包含项的个数
$LSp$	对向量 $LS$ 由大到小进行排序后得到的向量
$Su$	$X$ 的后缀项集概率权值矩阵,大小为 $m \times \max(LS)$
$Sup$	对 $Su$ 所有行向量按照由大到小排序后得到的矩阵

例如,在表 1 的示例中, $LT = [5, 3, 2, 3, 2, 4, 3, 4]$ 。当  $X = \{AC\}$  时, $XS = [1, 4, 6]$ , $LS = [3, 1, 2]$ , $LSp = [3, 2, 1]$ , $|X| = 2$ ,此时相应的  $Pu$ 、 $Su$  和  $Sup$  矩阵如表 3 所示。

表 3 当  $X = \{AC\}$  时表 1 数据库的变形

$XS$	$Pu$	$Su$	$Sup$	$LS$
1	[2.43, 3.95]	[7.44, 11.68, 10.2]	[11.68, 10.2, 7.44]	3
4	[15.84, 10.12]	[6.48]	[6.48]	1
6	[20.75, 15.47]	[20.64, 3.04]	[20.64, 3.04]	2

从表 3 可知,当  $X = \{AC\}$  时, $Pu(4, 1) = 15.84$ , $Su(6, 2) = 3.04$ , $Sup(1, 2) = 10.2$ 。在表 3 中没有显示的矩阵值都为 0,例如, $Pu(2, 2) = 0$ , $Su(4, 2) = 0$ , $Sup(4, 2) = 0$ 。

针对带有权值的不确定型事务数据库  $D$ ,给定一个模式指标  $PN$  和项集  $X$ ,该界值估算方法的基本

步骤如下:

**步骤 1** 对于一组给定的  $\alpha$  和  $\beta$ ,设计一个函数  $F(\alpha, \beta, Pu, Sup)$  满足以下条件:

$$PN(X', D_{X'}, H) \leq F(\alpha, \beta, Pu, Sup) \quad (5)$$

**步骤 2** 在合理的取值范围内搜索  $\alpha$  和  $\beta$  的取值,使函数  $F$  达到最大值, $\max_{\alpha, \beta} F(\alpha, \beta, Pu, Sup)$  就是  $X$  的超集关于  $PN$  的一个上界。

## 3 基于修正 BPSO 的模式指标上界求解方法

在上述基本框架中,核心问题是搜索  $\alpha$  和  $\beta$  的取值,使得函数  $F(\alpha, \beta, Pu, Sup)$  的取值达到最大。这是一个典型的组合优化问题,因此,可以采用 BPSO 算法来寻找最优解。

### 3.1 粒子编码

本文采用基于二进制值的粒子编码策略。给定一个模式指标  $PN$  和项集  $X$ ,为提高计算效率并减少所需存储空间,设粒子的维数等于  $N_1 + N_2$ ,其中, $N_1 = \lceil \lg |XS| \rceil$ , $N_2 = \lceil \lg \max(LS) \rceil$ 。种群中的第  $i$  个粒子  $C_i$  被编码为  $c_i^1 c_i^2$ ,其中, $c_i^1$  由  $N_1$  位二进制值构成,其相应的十进制数  $k_i^1$  表示  $\alpha$  的一个取值, $c_i^2$  由  $N_2$  位二进制值构成,其相应的十进制数  $k_i^2$  表示  $\beta$  的一个取值。例如,当  $X = \{AC\}$  时,根据表 1 示例可得  $N_1 = N_2 = 2$ ,此时粒子的编码方式与  $\alpha$  和  $\beta$  的映射关系如表 4 所示。

表 4  $C_i$  编码方式与  $\alpha$  和  $\beta$  的映射关系

$C_i$	$\alpha$	$\beta$
1010	2	2
0110	1	2
0111	1	3
1101	3	1

### 3.2 适应度函数

针对项集  $X$ 、 $\alpha$  和  $\beta$ ,适应度值为  $F(\alpha, \beta, Pu, Sup)$ ,具体计算步骤如下:

**步骤 1** 当  $LS(i) < \beta (1 \leq i \leq |LS|)$  时,从向量  $XS$  中剔除相应的  $XS(i)$  得到  $XS_\beta$ ,从而确定可能包含  $X'$  的候选事务集  $C$ (由于  $X'$  相对于  $X$  的扩展长度为  $\beta$ ,因此当  $X$  在其支持事务上的后缀长度小于  $\beta$  时,该事务中一定不包含  $X'$ )。

**步骤 2** 根据候选事务集  $C$  和  $XS_\beta$  中存储的事务编号,将  $Su(XS_\beta(j), \max(LS)) (1 \leq j \leq |XS_\beta|)$  中的元素按由大到小的顺序进行排序得到  $Sup(XS_\beta(j), \max(LS))$ ,进而对  $Sup(XS_\beta(j), \max(LS))$  中前  $\beta$  个元素的模式指标值进行求和得到  $sum_\beta$ 。

**步骤 3** 对  $Pu(XS_\beta(j), \max(LS))$  中前  $|X|$  个元素的模式指标值进行求和得到  $sum_{|X|}$ ,进而得到局部上界值  $sum_\beta + sum_{|X|}$ 。

**步骤 4** 从步骤 3 中得到的所有局部上界值中选出前  $\alpha$  个最大值,对它们进行求和后得到函数  $F$  的值。

**3.3 修正策略**

本文中粒子的速度和位置按照式(1)和式(2)进行更新。但是由于在函数  $F$  的计算过程中, $\alpha$  和  $\beta$  不是 2 个相互独立的参数,因此更新后可能会产生没有实际意义的粒子。例如,在表 1 中,当  $\alpha=2$  时, $LSp(2)=2$ ,即当  $X'$  可能被包含于 2 个事务时,其扩展长度一定不能超过 2,而在粒子的随机更新过程中,很可能会产生  $\alpha=2$  且  $\beta=3$  的粒子。为消除这些没有意义的粒子,本文提出一种粒子编码修正策略,该策略主要分为 2 个步骤:

**步骤 1** 当  $k_i^l \in \{1, 2, \dots, |XS|\}$  时, $\alpha$  的取值合法,转到步骤 2;否则,将粒子  $i$  在历代中最优值  $p_i$  的前  $N$  位赋值给  $c_i^l$ 。

**步骤 2** 当  $k_i^l \in \{0, 1, \dots, LSp(k_i^l)\}$  时,粒子合法,不需要修正;否则,将粒子  $i$  在历代中最优值  $p_i$  的后  $N$  位赋值给  $c_i^l$ 。

从上述步骤可以看出,所提修正策略的主要设计思想是当检测到无意义的粒子时,用个体最优解代替粒子中无意义部分的取值。采用这种修正策略,一方面可以保证种群中的所有粒子都具有实际意义,另一方面可以加快搜索速度,减少运行时间。

**3.4 算法流程**

重复执行第 3.1 节 ~ 第 3.3 节的过程,直至满足终止条件。基于修正 BPSO 的模式指标上界值估算方法的流程如图 1 所示。

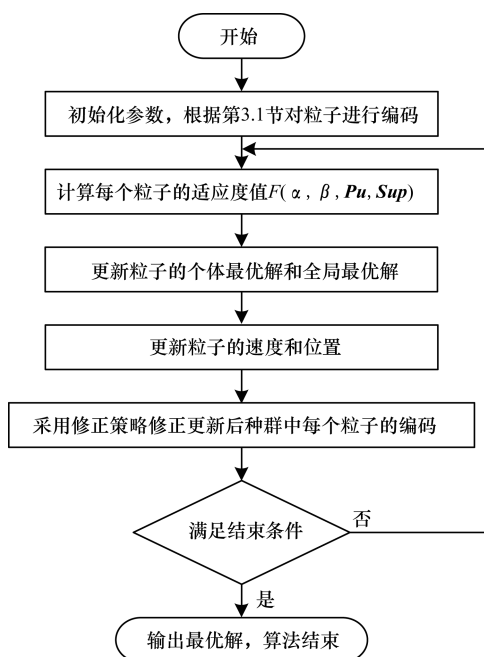


图 1 模式指标上界值估算方法流程

**4 实验验证**

模式效用<sup>[13-14]</sup>是一种常用的模式指标,本文在实验中以模式效用为例,验证所提通用模式指标上界估算方法针对带有权值的不确定型事务数据库的有效性。实验方案分为 2 个步骤:

1) 分别采用项集的事务加权效用值(Transaction Weighted Utilization, TWU)、本文方法所得的上界估算值和实际上界值(项集  $X$  所有超集的模式指标值中的最大值),来减少高效用模式挖掘算法在带有权值的不确定型事务数据库上所产生的候选项集数量。

2) 在上述 3 种界值计算方法下对比方法的运行时间和内存占用情况。由于目前针对不确定型数据库的高效用模式挖掘算法主要有 UHUI-Apriori<sup>[15]</sup>、PHUI-UP 和 PHUI-List<sup>[16]</sup>等,且都需要估算模式效用在不确定型事务数据库上的界值,即这些算法的实验过程和结果都很相似,因此,下文只展示了 UHUI-Apriori 算法与本文实验方案结合后的实验结果。

本节中所有算法都利用 Matlab 实现,并运行在一台装有 32 位 Windows7 操作系统的台式计算机上,该计算机带有 2 个英特尔酷睿 i5-4590 处理器和 4 GB 的 RAM。

**4.1 实验数据集**

本节实验运行在人工数据集 T10I4D100K 和实际数据集 mushroom、connect 和 accidents 上。这些都是关于高效用模式挖掘问题的常用数据集,其参数和特征分别如表 5、表 6 所示。

表 5 数据集参数说明

参数	含义
$\# D $	总事务个数
$\# I $	所包含项的总个数
AvgLen	平均事务长度
Type	数据类型(稀疏或密集)

表 6 各数据集特征

数据集名称	$\# D $	$\# I $	AvgLen	Type
T10I4D100K	100 000	870	10.1	稀疏
mushroom	8 124	119	7.2	密集
connect	67 557	132	43.0	密集
accidents	340 183	468	33.8	密集

为满足 UHUI-Apriori 算法对数据集的要求,根据文献[17]所提出的仿真模型生成分别在区间  $[1, 1\ 000]$  和  $[0, 1]$  中的服从对数正态分布的随机数,用来模拟这些数据集中效用值和存在概率。本文方法中所采用的概率权值对应于 UHUI-Apriori 算法中每个项的效用值与其相应存在概率的乘积。

### 4.2 实验结果与分析

在实验中, BPSO 算法的一般参数设置为:  $\omega = 0.9, c_1 = c_2 = 2$ , 种群规模  $popsiz = 30$ 。在该参数设置下, 比较 UHUI-Apriori 算法分别采用上述 3 种上界值后的候选项集数量, 结果如图 2 所示。从图 2 可以看出, UHUI-Apriori 算法采用实际上界值时生成的候选项集数量最少, 本文方法

次之, TWU 生成了最多的候选项集。该结果表明, 本文方法所得上界值比 TWU 所得上界值更接近于实际上界值。为进一步验证本文方法的有效性, 分析 UHUI-Apriori 算法在分别采用上述 3 种上界值后的运行时间和内存使用情况, 结果分别如图 3、图 4 所示。

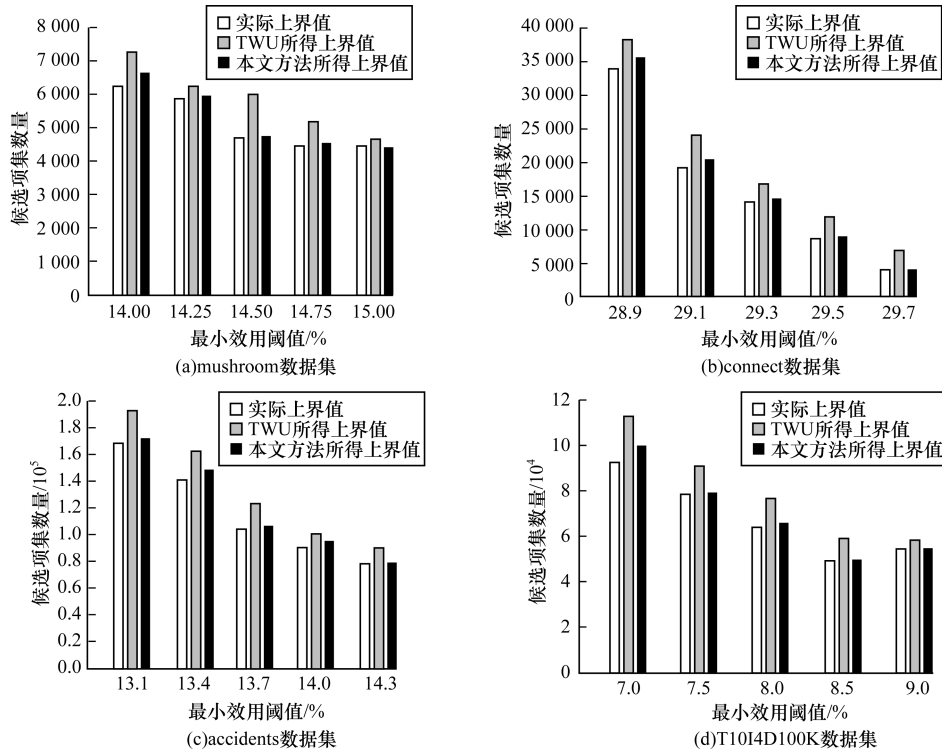


图 2 采用不同上界值估算方法时候选项集数量对比

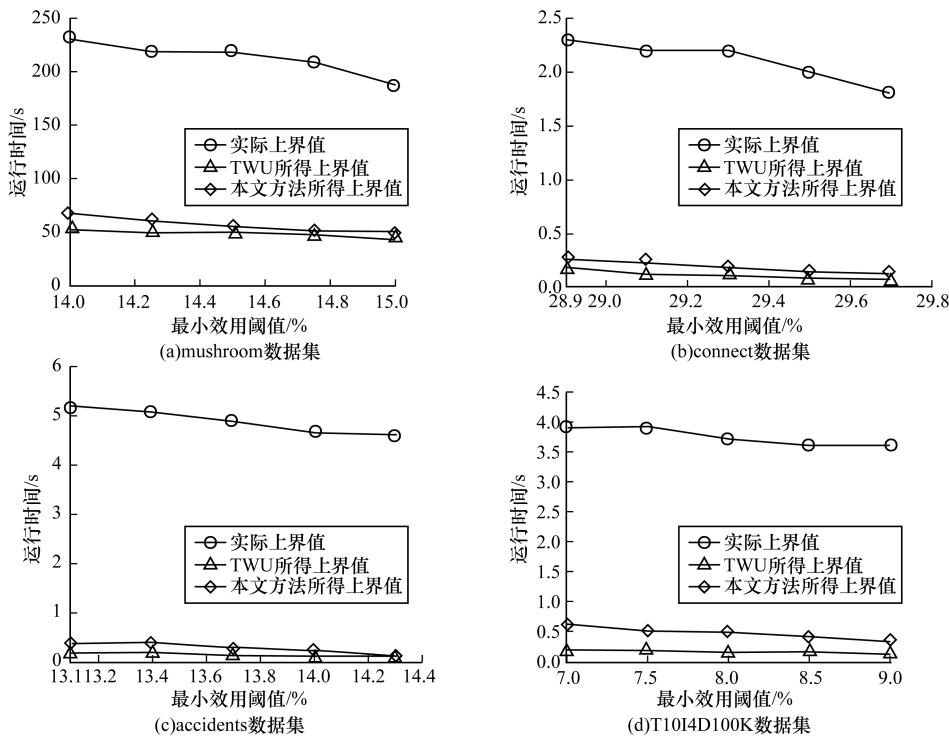


图 3 采用不同上界值估算方法时 UHUI-Apriori 运行时间对比

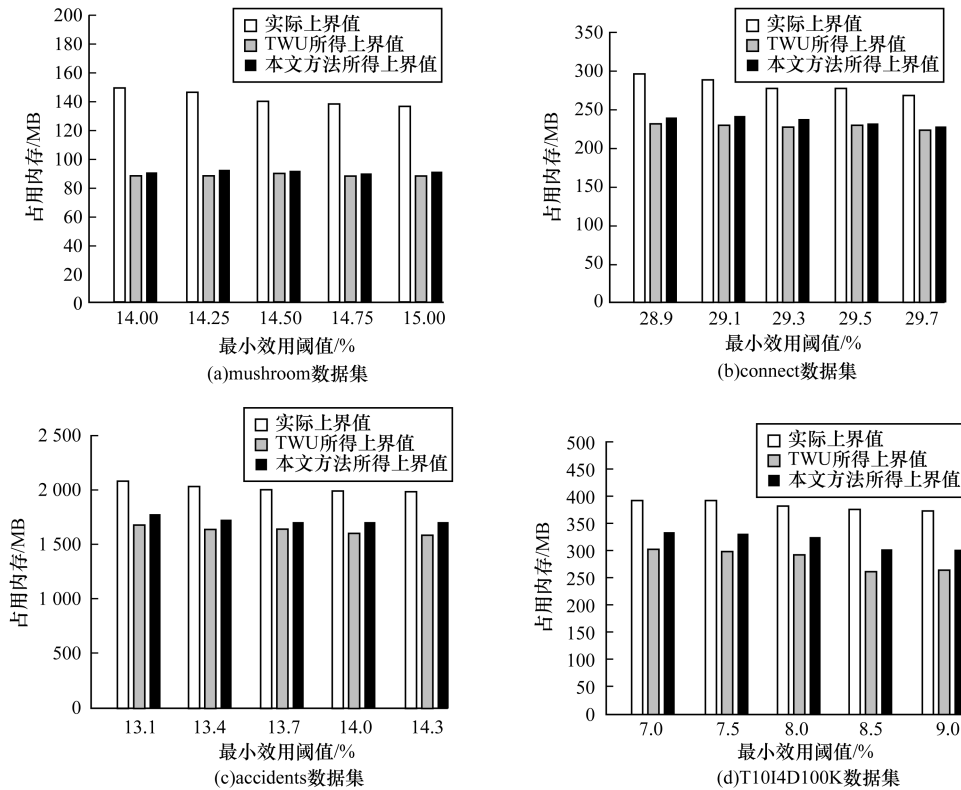


图4 采用不同上界值估算方法时 UHUI-Apriori 所占内存对比

从图3、图4可以看出,相比于采用TWU和本文方法估算所得上界值,当UHUI-Apriori算法采用实际上界值时,需要占用大量的运行时间和内存。这是因为在实际上界值的估算过程中,需要计算项集 $X$ 所有超集的模式效用值,因此存在组合爆炸问题,导致计算量巨大。虽然本文方法所生成的候选项集数量较小,但是当UHUI-Apriori算法采用本文方法所得上界值时,所需要的运行时间和内存都比TWU略大,尤其是在稀疏数据集T10I4D100K上差距明显。造成这种结果的原因主要为:1)TWU是通过事务的权值来松散地估算项集的权值,计算复杂度较小;2)本文界值估算方法是一种通用方法,不只适用于高效用模式的上界估计,因此相比于TWU,本文算法有较大的计算复杂度;3)由于本文界值估算方法需要对种群进行迭代更新,因此占用了部分运行时间与内存。

通过上述实验结果分析可以得出结论:本文上界估算方法省去了针对不同难约束模式指标都需要估算其相应上界的过程,且所估算出的上界值较紧凑,但代价是相对于特定的上界值估算方法可能会占用略多的运行时间和内存。但是,相对于计算项集的实际值,本文方法仍然节省了大量的运行时间和内存。

## 5 结束语

本文针对带有权值的不确定型事务数据库,提出一种基于修正BPSO的通用模式指标上界估算方法。该方法建立对难约束指标进行上界估算的统一框架,并基于修正BPSO算法实现对模式指标上界值的快速求解。通过对比UHUI-Apriori算法分别与TWU、本文方法估算所得上界值和实际上界值结合后的候选项集数量、运行时间和内存占用情况,验证了本文方法的可行性和高效性。下一步考虑将不确定性理论、智能搜索算法与通用上(下)界值估算方法进行深入结合,以提高本文方法的灵活性与通用性。

## 参考文献

- [1] ZHU F J, ZHANG Z, QU Q. A direct mining approach to efficient constrained graph pattern discovery [C]// Proceedings of SIGMOD'13. Washington D. C., USA: IEEE Press, 2013: 821-832.
- [2] GUNS T, NIJSSEN S, RAEDT L D. K-pattern set mining under constraints [J]. IEEE Transactions on Knowledge and Data Engineering, 2013, 25(2): 402-418.
- [3] AGRAWAL R, IMIELINSKI T, SWAMI A. Mining association rules between sets of items in large databases [C]// Proceedings of SIGMOD'93. Washington D. C., USA: IEEE Press, 1993: 207-216.
- [4] CAO L. Domain-driven data mining: challenges and prospects [J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(6): 755-769.

- [5] LIU Y, CHEN L, PEI J, et al. Mining frequent trajectory patterns for activity monitoring using radio frequency tag arrays[C]//Proceedings of IEEE International Conference on Pervasive Computing and Communications. Washington D. C., USA: IEEE Press, 2007: 37-46.
- [6] HAN J, PEI J, YIN Y. Mining frequent patterns without candidate generation [C]//Proceedings of ACM SIGMOD International Conference on Management of Data. New York, USA: ACM Press, 2000: 1-12.
- [7] YAO H, HAMILTON H J, BUTZ C J. A foundational approach to mining itemset utilities from databases [EB/OL]. [2017-07-05]. <http://www2.cs.uregina.ca/~butz/publications/siam04.pdf>.
- [8] WU C W, SHIE B E, TSENG V S, et al. Mining top-K high utility itemsets [C]//Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM Press, 2012: 78-86.
- [9] TANG L, ZHANG L, LUO P, et al. Incorporating occupancy into frequent pattern mining for high quality pattern recommendation [C]//Proceedings of the 21st ACM International Conference on Information and Knowledge Management. New York, USA: ACM Press, 2012: 75-84.
- [10] YUN U. Efficient mining of weighted interesting patterns with a strong weight and/or support affinity [J]. Information Sciences, 2007, 177(17): 3477-3499.
- [11] 张磊. 基于约束的频繁模式挖掘方法 [D]. 合肥: 中国科学技术大学, 2014.
- [12] 吴正佳, 罗月胜, 周玉琼, 等. 一种求解典型 JSP 的改进离散粒子群优化算法 [J]. 计算机应用研究, 2013, 30(8): 2406-2409.
- [13] LIN C W, HONG T P, LU W H. An effective tree structure for mining high utility itemsets [J]. Expert Systems with Applications, 2011, 38(6): 7419-7424.
- [14] LIU M, QU J. Mining high utility itemsets without candidate generation [C]//Proceedings of ACM International Conference on Information and Knowledge Management. New York, USA: ACM Press, 2012: 55-64.
- [15] LAN Y, WANG Y N, WANG Y, et al. Mining high utility itemsets over uncertain databases [C]//Proceedings of International Conference on Cyber-enabled Distributed Computing and Knowledge Discovery. Washington D. C., USA: IEEE Press, 2015: 235-238.
- [16] LIN C W, GAN W, FOURNIER-VIGER P, et al. Efficient algorithms for mining high-utility itemsets in uncertain databases [J]. Knowledge-Based Systems, 2016, 96 (C): 171-187.
- [17] LIU Y, LIAO W, CHOUDHARY A. A two-phase algorithm for fast discovery of high utility itemsets [C]//Proceedings of Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining. Berlin, Germany: Springer, 2005: 689-695.

编辑 吴云芳

(上接第 159 页)

- [4] CAN Z, DEMIRBAS M. A survey on in-network querying and tracking services for wireless sensor networks [J]. Ad Hoc Networks, 2013, 11(1): 596-610.
- [5] 高峰. 2.4 GHz 无线网络在视频监控中的应用 [J]. 铁道通信信号, 2011, 47(2): 61-63.
- [6] 李萌, 孙恩昌, 张延华. 无线信道模型研究与展望 [J]. 中国电子科学研究院学报, 2012, 7(4): 362-364.
- [7] WYNE S, SINGH A P, TUFVESSON F, et al. A statistical model for indoor office wireless sensor channels [J]. IEEE Transactions on Wireless Communications, 2009, 8(8): 4154-4164.
- [8] 黄少驰, 孔得强. 室内无线电波传播实验研究 [J]. 通信对抗, 2011, 5(1): 47-49.
- [9] 王萍, 勾天航, 李朋朋, 等. 室内走廊环境高频段宽带无线信道测量与建模 [J]. 电波科学学报, 2012, 27(3): 47-49.
- [10] 田勇, 唐祯安, 喻言. 室内无线传感器网络信道传输模型及统计分析 [J]. 控制与决策, 2014, 29(6): 1135-1138.
- [11] 朱莉, 顾能华, 姚英彪, 等. 基于 RSSI 的 WSN 二维对数搜索定位算法 [J]. 计算机工程, 2014, 40(4): 87-90.
- [12] 马健生, 金志强, 姜光, 等. 室内无线通信系统的路径损耗模型研究 [J]. 信息通讯, 2016(1): 213-216.
- [13] 徐琨, 刘宏立. 室内环境下无线传感器网络路径衰减特性 [J]. 传感器与微系统, 2016, 35(12): 11-13.
- [14] 张翔, 潘锦. 双频水平极化全向天线的设计 [J]. 电子科技, 2014, 25(5): 59-61.
- [15] 李冲. 收发天线高度对无线信道传播特性影响的测量与建模 [D]. 北京: 北京邮电大学, 2014.
- [16] 张飞, 耿红琴. 基于阴影衰落信道建模的 WSN 能量优化算法 [J]. 计算机工程, 2016, 42(6): 108-112.
- [17] MARTINEZ A, MOLINA J M, EGEE E. A accurate radio channel model for wireless sensor networks simulation [J]. Journal of Communications and Networks, 2005, 7(4): 78-85.
- [18] 沈杰, 姚道远, 黄河清, 等. 野外地表无线传感网信道传播模型的测定与分析 [J]. 光学精密工程, 2008, 16(1): 141-149.

编辑 顾逸斐