

引用格式:吴嵘,张姣玲,刘小兰.结合变异机制和量子 PSO 的关联规则挖掘算法[J].山东科技大学学报(自然科学版),2020,39(2):95-104.

WU Rong,ZHANG Jiaoling,LIU Xiaolan.An association rule mining algorithm based on mutation mechanism and QPSO[J].Journal of Shandong University of Science and Technology (Natural Science),2020,39(2):95-104.

## 结合变异机制和量子 PSO 的关联规则挖掘算法

吴 嵘<sup>1</sup>,张姣玲<sup>2</sup>,刘小兰<sup>3</sup>

(1.广东科贸职业学院 信息与自动化学院,广东 广州 510640;

2.广东技术师范大学 数学与系统科学学院,广东 广州 510665;3.华南理工大学 数学学院,广东 广州 510641)

**摘 要:**针对数据集中的关联规则挖掘问题,提出一种基于改进量子粒子群优化(improved quantum particle swarm optimization, IQPSO)算法的关联规则挖掘方法。首先,将数据实例以量子比特形式表示,构建一个基于量子进化算法(quantum evolutionary algorithm, QEA)的关联规则挖掘基础框架。然后,在该基础框架上,采用新的量子角度更新公式,即使用 QPSO 代替 QEA 实现关联规则挖掘。最后,为了进一步提高 QPSO 算法的收敛性能,融入变异机制和动态惯性权重对其进行改进,加快其收敛速度和跳出局部最优的能力。在 UCI 和课程成绩数据集上的实验结果表明,提出的算法能够快速且有效地挖掘出关联规则,相比其他几种算法,挖掘到的关联规则价值更高。

**关键词:**关联规则挖掘;量子粒子群优化;变异机制;动态惯性权重;量子进化算法

中图分类号:TP311

文献标志码:A

文章编号:1672-3767(2020)02-0095-10

DOI:10.16452/j.cnki.sdkjzk.2020.02.011

### An association rule mining algorithm based on mutation mechanism and QPSO

WU Rong<sup>1</sup>,ZHANG Jiaoling<sup>2</sup>,LIU Xiaolan<sup>3</sup>

(1.School of Information and Automation,Guangdong Polytechnic of Science and Trade,

Guangzhou,Guangdong 510640,China; 2.School of Mathematics and Systems Science,

Guangdong Polytechnic Normal University,Guangzhou,Guangdong 510665,China;

3.School of Mathematics,South China University of Technology,Guangzhou,Guangdong 510641,China)

**Abstract:** In view of the existing problems of the association rule mining in big data set, an association rule mining method based on improved quantum particle swarm optimization (IQPSO) algorithm was proposed in this paper. Firstly, the data instances were represented in the form of quantum bits, and a basic framework of association rule mining based on quantum evolutionary algorithm (QEA) was constructed. Then, QPSO was used instead of QEA to construct a new quantum angle updating formula to realize association rule mining. Finally, mutation mechanism and dynamic inertia weight were integrated to improve QPSO so as to accelerate convergence speed and the ability to jump out of local optimum. The experimental results on UCI and student test score data sets show that the proposed method can mine the association rules quickly and effectively and the fitness value of association rules mined by this algorithm is higher than those mined by other algorithms.

**Key words:** association rule mining; quantum-behaved particle swarm optimization; mutation mechanism; dynamic

收稿时间:2019-01-24

基金项目:国家自然科学基金项目(11801097)

作者简介:吴 嵘(1976—),女,广东韶关人,实验师,硕士,主要研究领域为数据挖掘等。

刘小兰(1979—),女,江西安福人,副教授,博士,主要从事智能优化算法、机器学习方面的研究,本文通信作者。

E-mail:serwu1@126.com

inertia weight; quantum evolutionary algorithm

关联规则挖掘是一种用来查找数据库中的频繁项或属性集之间相关性和因果关联的数据挖掘方法<sup>[1]</sup>。从数据库中执行关联规则挖掘是一个 NP 完全问题,其取决于数据库、机器学习和优化算法的研究。

目前已有的多种关联规则挖掘算法可以分为两类:一是精确算法<sup>[2]</sup>,这类算法一定程度上可以获得最优解,但需要多次扫描数据库,产生大量候选项集,对候选项集进行模式匹配时需花费大量时间<sup>[3]</sup>,如 Apriori 和 FP-Growth 算法。二是智能进化算法<sup>[4]</sup>,这类算法也能够给出很好的解,且执行时间会明显缩短,例如遗传算法、模拟退火算法等。

大型数据库中的关联规则挖掘是一个复杂的过程,使用精确算法非常昂贵,而智能进化算法在这个领域提供了很大的帮助。其中量子进化算法 (quantum-inspired evolutionary algorithm, QEA)<sup>[5]</sup> 使用量子比特作为信息的最小单元<sup>[6]</sup> 进行概率表示。量子比特个体为量子比特组成的串,称为多量子比特。量子比特个体的优点是可以在搜索空间中以概率表示状态的线性叠加。因此,量子比特表示法比遗传算法中的染色体表示法具有更好的群体多样性特征<sup>[7]</sup>。QEA 中的量子门为一种变体运算符,用来使个体转向更好的解,并最终收敛到单一最优解<sup>[8]</sup>。量子粒子群优化 (quantum-behaved particle swarm optimization, QPSO)<sup>[9]</sup> 是 QEA 和粒子群优化 (PSO) 的混合算法,采用一种新的称为量子角度的量子比特表达机制。并利用 PSO 中的位置更新机制来更新 QEA 中的量子比特,使其寻优能力优于传统 QEA。

基于上述分析,提出一种基于改进 QPSO 的关联规则挖掘方法,并在 UCI 和课程成绩数据集上进行了实验验证,结果表明,提出的方法在执行时间和挖掘规则质量上都具有一定的优势。提出方法的主要创新点在于:将数据实例以量子比特形式表示,构建一个基于 QEA 的关联规则挖掘基础框架;使用 QPSO 代替 QEA 实现关联规则挖掘,并融入变异机制和动态惯性权重形成 IQPSO,以加快其收敛速度和跳出局部最优的能力。

## 1 关联规则挖掘问题描述

### 1.1 问题定义

关联规则挖掘中,令  $I = \{i_1, i_2, \dots, i_m\}$  为一组布尔型属性,称为项目;  $S = \{s_1, s_2, \dots, s_n\}$  为多组数据实例记录,其中每个数据实例  $s_i \in S$  由  $I$  中的不同属性构成。数据实例  $s_i$  中若存在某个布尔属性,则意味着其值为 1,如果不存在,则其值为 0。例如,令  $I = \{A, B, C\}$  为一组布尔型属性,  $S = \{\langle A, B \rangle, \langle C \rangle, \langle C \rangle\}$  为多组数据实例,那么多组  $S$  可以重写为:

$$S = \{\langle A=1, B=1, C=1 \rangle, \langle A=0, B=0, C=1 \rangle, \langle A=0, B=0, C=1 \rangle\}。 \quad (1)$$

一个关联规则可表示为:如果  $C$  成立,则为  $P$ 。其中  $C$  为条件,  $P$  为预测,  $C, P \subset 1$  并且  $C \cap P = \emptyset$ 。

本研究连接关联规则,其中  $C$  是一个或多个条件的连接,  $P$  是一个或多个预测的连接。文中所涉及的符号有:  $|C|$ —规则  $C$  包含的数据实例数量;  $|P|$ —规则  $P$  包含的数据实例数量;  $|C \& P|$ —规则  $C$  和  $P$  同时包含的数据实例数量;  $N$ —要挖掘的数据实例总数。

规则的置信度  $b$  表示在知道  $C$  被观察时  $P$  出现的概率,  $b = \frac{|C \& P|}{|C|}$ 。预测频率  $a = \frac{|P|}{N}$ , 支持度 =  $\frac{|C \& P|}{|N|}$ 。

### 1.2 规则适应度函数

关联规则的性能通常由置信度和支持度表示,为了更好地量化规则质量,采用文献[10]提出的适应度函数  $F$  来评估候选规则的质量。定义规则支持度为满足规则中  $C$  的数据实例的百分比,置信度  $b$  为满足规则 (IF  $C$ , THEN  $P$ ) 的数据实例的百分比。该适应度函数是基于信息论推导的,综合考虑了置信度和支持度。定义一个 J-度量  $J_m$ , 计算表达式为:

$$J_m = \frac{|C|}{N} \times \left( b \times \log \left( \frac{b}{a} \right) \right), \quad (2)$$

适应度函数

$$F = \frac{w_1 \times J_m + w_2 \times \frac{n_{pu}}{n_r}}{w_1 + w_2} \quad (3)$$

式中:  $n_{pu}$  为潜在有用属性的数量。如果至少存在一个数据实例在  $C$  部分和预测属性中具有指定的  $A$  值, 则认为该属性  $A$  是潜在有用的;  $n_r$  为规则中  $C$  包含的属性总数;  $w_1$ 、 $w_2$  为用户自定义权重, 这里分别设置为 0.6 和 0.4;  $F$  的取值在 0~1 之间, 值越大说明规则的质量越好。

## 2 基于 QEA 的关联规则挖掘

量子进化算法(QEA)与经典进化算法(EA)类似, 只是采用了量子个体来代替传统个体进行迭代。量子个体中使用基于量子比特的编码方式<sup>[11-12]</sup>, 即用一对复数定义一个量子比特位。这种表示方法可以表征任意的线性叠加状态。另外, 在种群更新中, 根据量子的叠加特性和变迁理论, 通过量子门变换来产生新的个体。在数据集中应用 QEA 进行规则挖掘时, 必须将数据实例通过量子来表示。基于 QEA 算法的关联规则挖掘过程如算法 1 描述。

### 算法 1: QEA 关联规则挖掘算法

输入: 实例数据集, 最小置信度, 支持度阈值, 种群数量, 最大迭代次数

输出: 一组最佳关联规则

**begin**

$t \leftarrow 0$ ;

初始化包含多个量子比特个体  $Q(t)$  的种群;

把  $Q(t)$  投影到二进制解  $P(t)$  中;

计算  $P(t)$  的适应度函数;

**If** 存在关联规则

**Then** 从每个  $P(t)$  上生成关联规则;

**end if**

存储  $P(t)$  中的最佳解;

**while**(结束条件) **do**

$t \leftarrow t + 1$ ;

    把  $Q(t-1)$  投影到二进制解  $P(t)$  中;

    计算  $P(t)$  的适应度函数;

**If** 存在关联规则

**Then** 从每个  $P(t)$  上生成关联规则;

    使用量子门更新  $Q(t)$ ;

**end if**

    存储  $P(t)$  中的最佳解;

**end do**

**end**

QEA 算法主要包含 4 个部分:

1) 初始化量子比特个体  $Q(t)$ : 其中  $Q(t) = \{q_1^t, q_2^t, \dots, q_n^t\}$ ,  $t$  表示迭代次数,  $n$  表示种群数量。另外,  $\alpha_i$  和  $\beta_i$  的值都初始化为  $1/\sqrt{2}$ 。

2)  $Q(t)$  投影到二进制解  $P(t)$ : 通过观察种群  $Q(t)$  的状态, 生成二进制解  $P(t) = \{x_1^t, x_2^t, \dots, x_n^t\}$ ,  $x_n^t = (x_1, x_2, \dots, x_m)$ 。对于多量子比特中的每一比特  $x_i (i \in m)$  通过一个 0 和 1 之间的随机变量来赋值: 如果  $rand(0, 1) > |\beta_i|^2$ , 则生成 1; 否则生成 0。即表示为:

$$x_i = \begin{cases} 1, & \text{if } \text{rand}(0,1) > |\beta_i|^2 \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

3) 计算  $P(t)$  的适应度函数:通过适应度函数  $F$  对每个二进制解  $P(t)$  进行评估。

4) 使用量子门更新  $Q(t)$ :通过量子旋转门来更新个体角增量。

量子门  $U(\Delta\theta)$  是一个可变运算符,可以根据具体问题进行选择。本研究使用的量子门定义为:

$$U(\Delta\theta_i) = \begin{bmatrix} \cos(\xi(\Delta\theta_i)) & -\sin(\xi(\Delta\theta_i)) \\ \sin(\xi(\Delta\theta_i)) & \cos(\xi(\Delta\theta_i)) \end{bmatrix}, \quad (5)$$

式中:  $\xi(\Delta\theta_i) = s(\alpha_i, \beta_i) \times \Delta\theta_i$ ;  $s(\alpha_i, \beta_i)$  和  $\Delta\theta_i$  分别表示旋转方向和角度。

### 3 基于改进 QPSO 的关联规则挖掘

#### 3.1 QPSO 算法

由于传统 QEA 中通过量子旋转门来更新量子角增量,操作复杂且更新角度固定,使其容易陷入局部最优<sup>[13]</sup>。为此, Sun 等<sup>[14]</sup> 引入了 PSO 算法中的位置更新公式替代 QEA 中的量子旋转门来更新角增量,形成 QPSO 算法。

为了引入 PSO 位置更新策略,需定义一个量子角度  $\theta$ , 那么量子比特表示为  $[\theta]$ , 其等价于原始的量子比特  $\left[ \frac{\sin(\theta)}{\cos(\theta)} \right]$ , 满足条件:

$$|\sin(\theta)|^2 + |\cos(\theta)|^2 = 1. \quad (6)$$

那么, 可以将多量子比特  $\begin{bmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_m \\ \beta_1 & \beta_2 & \dots & \beta_m \end{bmatrix}$  替换为:  $[\theta_1 | \theta_2 | \dots | \theta_m]$ 。

量子旋转门变为:

$$[\alpha'_i \quad \beta'_i]^T = U(\Delta\theta_i) [\alpha_i \quad \beta_i]^T. \quad (7)$$

其中,  $U(\Delta\theta_i) = \begin{bmatrix} \cos(\xi(\Delta\theta_i)) & -\sin(\xi(\Delta\theta_i)) \\ \sin(\xi(\Delta\theta_i)) & \cos(\xi(\Delta\theta_i)) \end{bmatrix}$  通过  $[\theta'_i] = [\theta_i + \xi(\Delta\theta_i)]$  来代替。

QPSO 使用 PSO 的群智能概念, 将群体中的所有多量子比特视为智能种群, 即量子群。首先, QPSO 找到局部最佳量子角, 并从局部最佳量子角中找到全局最佳量子角。然后根据这些值, 用量子角更新公式更新量子角。

基于 QEA 的 QPSO 步骤如下:

1) 使用量子角编码量子比特  $Q(t)$ :  $Q(t) = \{q_1^t, q_2^t, \dots, q_m^t\}$ ,  $q_i^t = [\theta_{j1}^t | \theta_{j2}^t | \dots | \theta_{jm}^t]$ 。

2) 通过  $|\cos(\theta)|^2$  观察  $Q(t)$  的状态, 把  $Q(t)$  投影到二进制解  $P(t)$ 。对于多量子比特中的每一比特  $x_j$ , 其由量子角  $\theta$  与 0 和 1 之间的随机变量得到: 如果  $\text{rand}(0,1) > |\cos(\theta)|^2$ , 则生成 1; 否则生成 0, 即:

$$x_i = \begin{cases} 1, & \text{if } \text{rand}(0,1) > |\cos(\theta)|^2 \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

3) 使用以下 PSO 位置更新公式替换传统“量子门更新  $Q(t)$ ”步骤:

$$\theta_{ji}^{t+1} = \omega \times \theta_{ji}^t + c_1 \times \text{rand}() \times (\theta_{ji}^t(p_{\text{best}}) - \theta_{ji}^t) + c_2 \times \text{rand}() \times (\theta_{ji}^t(g_{\text{best}}) - \theta_{ji}^t), \quad (9)$$

其中:  $\theta_{ji}^{t+1}$ 、 $\theta_{ji}^t$ 、 $\theta_{ji}^t(p_{\text{best}})$  和  $\theta_{ji}^t(g_{\text{best}})$  分别为第  $i$  个多量子比特中的第  $i$  个量子比特的下一步角度、当前角度、个体最佳和全局最佳;  $c_1$  和  $c_2$  为学习因子, 这里分别设定为 1.42 和 1.57;  $\omega$  为惯性权重, 通常设定为固定值。

#### 3.2 改进 QPSO 算法(IQPSO)

为了提高 QPSO 算法的收敛性和跳出局部最优的能力, 分别引入动态惯性权重和变异操作对其进行改进。

##### 3.2.1 动态惯性权重

从式(9)可以看出, 惯性权重  $\omega$  决定了对当前量子角度的继承程度, 传统 PSO 中将其设置为恒定值。

然而,恒定的  $\omega$  不能适应动态的收敛过程。在算法搜索的前期,较大的  $\omega$  有利于跳出局部最优;在搜索后期,较小的  $\omega$  有利于局部寻优<sup>[15]</sup>。为此,引入了动态惯性权重,使  $\omega$  随着迭代数量的增加而递减,表示为:

$$\omega = \omega_{\min} + \frac{\omega_{\max} - \omega_{\min}}{T_{\max}} \times (T_{\max} - ite), \quad (10)$$

式中: $\omega_{\max}$  和  $\omega_{\min}$  为  $\omega$  的上限值和下限值,通常设定为 0.9 和 0.4; $T_{\max}$  为算法的最大迭代轮数, $ite$  为当前迭代轮数。

### 3.2.2 变异操作

QPSO 算法虽然引入了 PSO 中的角度更新机制和动态惯性权重,但其在迭代后期的角度变化幅度越来越小,仍有可能陷入局部最优。为此,本研究融入了一个变异机制来改进 QPSO 算法:当判断其迭代解停止不前时,启动变异过程。

首先,建立一种机制来判断 QPSO 寻优过程是否处于停止不前状态。本研究以全局最佳值  $F_{g\text{best}}$  与平均个体最佳值  $F_{p\text{best}}$  的比例关系  $k_{g-p}$  作为判断依据,其中设定取  $N=20$  次连续迭代中的个体最佳值来取平均。表达式如下:

$$k_{g-p} = F_{g\text{best}} / \sum_{n=1}^N F_{p\text{best}_n} \quad (11)$$

如果连续  $N$  次迭代后,量子的个体最佳角度几乎没有改变,即  $k_{g-p}$  趋于 1,那么判断此时的寻优过程处于停止不前状态。

然后,建立一种变异机制。选择一些量子进行变异操作,根据粒子与全局最优角度的距离作为变异概率,将远离最优角度的量子进行大概率变异,以提高种群跳出局部最优的能力。其中,距离用适应度值来表示。假设量子种群数量为  $M$ ,那么各量子的变异概率

$$p_i = F / \sum_{i=1}^M F_i \quad (12)$$

最后,通过随机函数  $\text{rand}(0,1)$  为每个量子产生一个随机值,若该值小于该量子的变异概率  $p_i$ ,则通过量子非门实现量子中  $1/2$  个量子位的变异,量子非门变异操作表示如下:

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \cos\theta_i \\ \sin\theta_i \end{bmatrix} = \begin{bmatrix} \sin\theta_i \\ \cos\theta_i \end{bmatrix} \quad (13)$$

### 3.3 关联规则挖掘过程

基于上述 IQPSO 算法,最终得到关联规则挖掘过程如算法 2 所示。

#### 算法 2: IQPSO 关联规则挖掘算法

输入:实例数据集,最小置信度,支持度阈值,种群数量  $M$ ,最大迭代次数  $T_{\max}$ ,学习因子,权重  $\omega$  的上限和下限值,变异概率  $p_i$

输出:一组最佳关联规则

begin

$t \leftarrow 0$ ;

初始化包含多个量子比特个体  $Q(t)$  的种群,并使用量子角编码量子比特;

把  $Q(t)$  投影到二进制解  $P(t)$  中;

计算  $P(t)$  的适应度函数;

If 存在关联规则

Then 从每个  $P(t)$  上生成关联规则;

end if

存储  $P(t)$  中的最佳解;

while(结束条件) do

$t \leftarrow t + 1$ ;

```

把  $Q(t-1)$  投影到二进制解  $P(t)$  中;
计算  $P(t)$  的适应度函数;
If 存在关联规则
    Then 从每个  $P(t)$  上生成关联规则;
end if
使用包含动态  $w$  的式(12)更新  $Q(t)$  中量子角;
存储  $P(t)$  中的最佳解;
If  $N$  次迭代后寻优过程处于停止不前状态
    Then 计算量子变异概率,通过量子非门进行变异操作;
end if
end do
end

```

## 4 实验及分析

### 4.1 实验设置

将提出的 IQPSO 关联规则挖掘算法与文献[16]提出的 Apriori 规则挖掘算法进行比较。为了体现 IQPSO 的改进效果,将其与传统 QPSO<sup>[9]</sup> 的关联规则算法进行比较。所有实验均在配备 Intel Core i7 处理器(3.4 GHz 主频、8 G 内存)和 Windows 10 的 PC 电脑上进行。使用 MATLAB 编程。其中, IQPSO 算法的种群数量为 30, 最大迭代次数设置为 200 次, 变异概率  $p_i = 0.01$ 。Apriori 和 QPSO 的参数设置分别参考文献[16]和文献[9]。

为了验证本文算法的有效性, 选择了一个由加州大学建立的国际标准机器学习数据库(UCI)中的幼儿园数据集作为实验数据集。幼儿园数据库最初是为幼儿园入学申请排名而设计的分层数据模型, 包含 12 960 个实例和 9 个属性, 以及各自的分类。另外, 为了将本算法应用到实际的高校教学管理系统中, 更好服务于课程设置和教学改革, 收集了学校几届计算机应用技术专业的 200 名大学生的 10 门课程成绩, 构建了一个课程成绩数据集, 按成绩分为优、良、中、及格和不及格 5 类。从学生各门课程的成绩中挖掘关联规则, 为学校合理安排课程和教师了解学生情况提供帮助。两个数据集的属性如表 1 所示。

表 1 数据库属性  
Tab. 1 Attributes of the database

UCI 幼儿园数据库			课程成绩数据集	
	属性	属性值	属性	属性值
1	父母	平凡, 良好, 优秀	离散数学(DM)	
2	幼儿园	适合, 不太适合, 不正当, 关键, 非常关键	操作系统(OS)	
3	形式	完全, 不完全, 培养	Unix 原理(UNIX)	
4	孩子	1, 2, 3, 更多	计算机维护与升级(CMU)	
5	住房	方便, 不方便, 关键	计算机网络(CN)	
6	金融	方便, 不方便	计算机基础(CCB)	优、良、中、及格和不及格
7	社会	没有问题, 小问题, 有问题	数据库设计(DBD)	
8	身体	健康, 非常健康, 不健康	微机原理(DEM)	
9	推荐	不推荐, 推荐, 强烈推荐, 优先, 特别优先	计算机硬件原理(CUE)	
10			电装实习(EE)	

### 4.2 算法收敛速度分析

为了验证 IQPSO 算法在收敛速度方面的优势, 采用标准的最优化问题测试函数-Sphere 函数, 该函数

是一个理论最小值为 0 的非线性对称的单峰函数,用于测试寻优速度,表达式为:

$$f(x) = \sum_{i=1}^n x_i^2, -100 \leq x \leq 100. \quad (14)$$

设置各种算法的种群数量为 30,最大迭代次数为 250,Sphere 函数的维度为 20。三种算法对 Sphere 函数进行极小值搜索的收敛曲线如图 1 所示。可以看到,三种算法都能收敛,但 IQPSO 算法的收敛速度最快,大约在 80 次迭代后就收敛到 0,而 PSO 和 QPSO 算法分别需要约 180 次和 120 次迭代。这是因为,IQPSO 中融入了变异机制和动态权重机制,这些操作的计算量很低,不会增加算法单次运行的时间。然而,这些改进可有效避免算法陷入局部最优,减少无效搜索的次数,提高算法的收敛速度。

#### 4.3 规则挖掘的执行时间比较

为了验证提出的算法在关联规则挖掘应用上的实际性能,在幼儿园数据库上进行了相关实验。在实例数目和置信度阈值一定时,对不同支持度阈值下的算法在运行时间方面进行比较。其中,选择 4 000 个实例进行实验,最小置信度为 0.7,支持度阈值为 0.5~0.75。当各种算法挖掘的规则达到最小支持度阈值时停止,并统计执行时间。实验结果如图 2 所示。

从图 2 可以看到,随着支持度阈值的增加,算法所运行的时间逐渐减少。这是因为当支持度阈值比较小时,能够满足条件的关联规则较多,算法的运行时间较长。其中,QPSO 算法的性能优于 Apriori 算法,这是因为 QPSO 算法是一种高效的智能优化算法,能够快速收敛到最优解。而 IQPSO 算法执行时间最短,且对于传统 QPSO 算法有明显提高。这是因为,IQPSO 中融入了变异机制和动态权重机制,提高算法的收敛速度,缩短了挖掘关联规则的执行时间。

接着,在固定最小置信度和最小支持度下,在不同数量的实例数据集上进行实验,比较各种算法的执行时间。其中,置信度设置为 0.7,最小支持度设置为 0.65,数据集实例数量为 4 000 到 8 000,结果如图 3 所示。可以看出,随着实例数目的增加,算法的运行时间均逐渐增加。在各种情况下,IQPSO 的运行时间都是最短的,且随着实例数量的增加,改善效果更加明显。

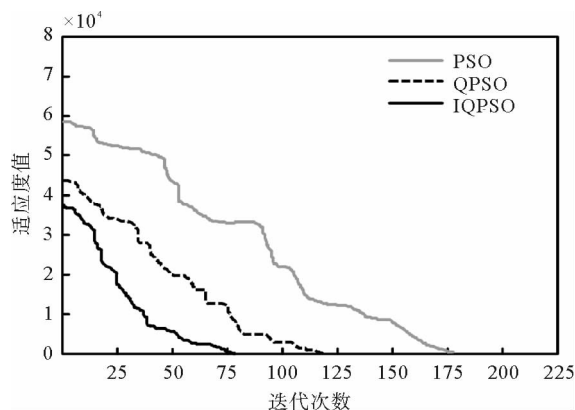


图 1 20 维 Sphere 函数的收敛曲线

Fig. 1 Convergence curve of 20-dimensional Sphere function

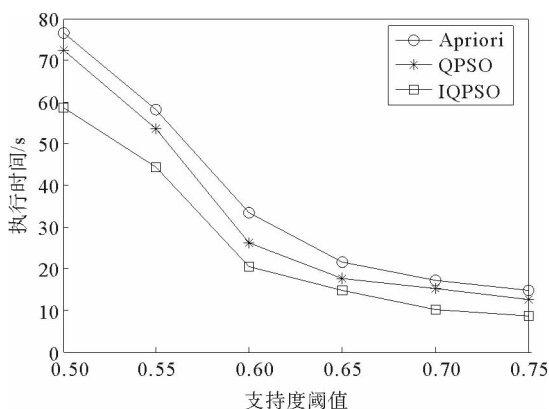


图 2 不同支持度阈值下各种算法的执行时间

Fig. 2 Execution time of each algorithm under different support thresholds

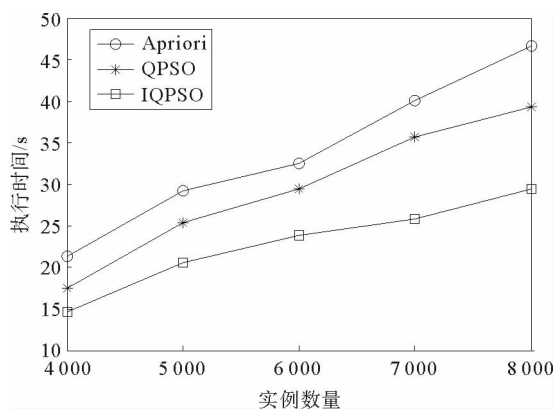


图 3 不同实例数量下各种算法的执行时间

Fig. 3 Execution time of each algorithm under different number of instances

另外,对于课程成绩数据集,由于实例数量较少,各种算法的运行时间都比较短(3 s 以内)且差距很小,

不像在幼儿园数据库上能看出明显差异。

#### 4.4 规则挖掘质量比较

从包含 4 000 个实例的幼儿园数据库中进行规则挖掘来比较所挖掘的规则质量。这里指定一个目标属性,即“推荐”,将 IQPSO 算法与文献[16]的 Apriori 算法和文献[17]基于遗传算法(GA)的挖掘算法进行对比。

对于目标“推荐=不推荐”,利用三种算法发现的最佳规则如表 2 所示。对比发现,相比其他两种算法,提出的 IQPSO 算法发现的规则更加有价值。例如:

IF “幼儿园=适合”、“孩子=2”、“住房=不方便”、“金融=不方便”、“社会=没有问题”并且“身体=不健康”;

THEN “推荐=不推荐”。

该规则的支持度 $|C \& P|$ 为 446,置信度 $b$ 为 0.97,适应度为 0.436。

对于目标“推荐=特别优先”,利用三种算法发现的最佳规则如表 3 所示。Apriori 规则的置信度 $b$ 为 0.94,适应度为 0.411。GA 规则的置信度和适应度要比 Apriori 规则高,因为 GA 是一种群体智能算法,收敛效果较好。类似表 2,除了 GA、Apriori 发现的规则外,IQPSO 算法还发现了其他更有趣的规则,如表 3,其支持度 $|C \& P|$ 为 348,置信度 $b$ 为 1.00,适应度=0.438。

上述两个规则挖掘实验结果表明了 IQPSO 关联规则挖掘算法所挖掘的规则具有更高的价值。这是因为 GA 算法的全局搜索能力较弱,容易陷入局部最优。IQPSO 算法通过融入变异机制和动态惯性权重,对 QPSO 进行了更深层次的改进,提高了算法性能,所以能够挖掘到最优的规则。

表 2 幼儿园数据库目标“推荐=不推荐”的规则挖掘结果

Tab. 2 Rule mining results with the goal of “recommendation=non-recommendation” in kindergarten database

算法	规则	支持度	置信度	适应度
Apriori <sup>[16]</sup>	IF“住房=不方便”并且“金融=不方便”;THEN“推荐=不推荐”	240	0.83	0.403
GA <sup>[17]</sup>	IF“孩子=2”、“住房=不方便”、“金融=不方便”并且“身体=不健康”;THEN“推荐=不推荐”	376	0.94	0.418
IQPSO	IF“幼儿园=适合”、“孩子=2”、“住房=不方便”、“金融=不方便”、“社会=没有问题”并且“身体=不健康”;THEN“推荐=不推荐”	446	0.97	0.436
	IF“住房=不方便”并且“身体=不健康”;THEN“推荐=不推荐”	351	0.92	0.410

表 3 幼儿园数据库目标“推荐=特别优先”的规则挖掘结果

Tab. 3 Rule mining results with the goal of “recommendation=special priority” in kindergarten database

算法	规则	支持度	置信度	适应度
Apriori <sup>[16]</sup>	IF“幼儿园=非常关键”并且“身体=非常健康”;THEN“推荐=特别优先”	285	0.94	0.411
GA <sup>[17]</sup>	IF“父母=优秀”、“幼儿园=非常关键”、“住房=方便”并且“身体=非常健康”;THEN“推荐=特别优先”	323	0.96	0.425
IQPSO	IF“父母=优秀”、“幼儿园=非常关键”、“孩子=1”、“住房=方便”、“金融=方便”、“社会=小问题”并且“身体=非常健康”;THEN“推荐=特别优先”	<b>348</b>	<b>1.00</b>	<b>0.438</b>

然后,在学生成绩数据集中进行同样的实验,三种算法挖掘出来的前两个规则分别如表 4 所示。这里举例分析本方法挖掘出的第一条关联规则:

IF “EE>=良好”并且“CUE>=及格”;THEN “CMU>=中等”。

这条规则的意思是,一个学生的电装实习(EE)成绩优良且计算机硬件原理(CUE)及格,通常计算机维护与升级(CMU)的成绩也会中等以上。这就说明电装实习与计算机维护与升级之间存在很大的相关性,电装实习优秀的学生通常动手能力较好,加上一定的硬件原理基础,就会使其在计算机维护与升级中获得好成绩。所以在今后教学中可以加强对对学生电路组装等方面动手能力提升的关注。



表 4 可见,在三种算法中,IQPSO 算法获得规则的适应度最高,说明了本文方法的有效性。

表 4 学生成绩数据集中规则挖掘结果举例  
Tab. 4 Examples of rule mining results in student test score database

算法	规则	支持度	置信度	适应度
Apriori <sup>[16]</sup>	IF“EE>=中等”并且“UNIX=不及格”;THEN“CN=不及格”	337	0.92	0.421
	IF“OS=优秀”并且“UNIX=不及格”;THEN“CN=不及格”	304	0.88	0.382
GA <sup>[17]</sup>	IF“Dm>=良好”并且“UNIX=不及格”;THEN“CN=不及格”	336	0.94	0.425
	IF“CCB>=良好”并且“CN>=及格”;THEN“CMU>=中等”	351	0.90	0.405
IQPSO	IF“EE>=良好”并且“CUE>=及格”;THEN“CMU>=中等”	355	0.94	0.438
	IF“CUE=不及格”并且“CN>=及格”;THEN“CMU=不及格”	348	0.92	0.426

在关联规则挖掘应用中,算法会挖掘出很多条规则,通常只需要选取一些高质量的规则来分析和指导实践。为了更加全面地评测各种算法的性能,通过上文描述的适应度函数来表征规则质量。在 2 个数据集上分别选择出前 10 条高质量规则,然后计算平均适应度值,结果如表 5 所示。从表 5 可以看出,IQPSO 算法不仅能够挖掘出最优关联规则,在前 10 条规则的平均性能方面也具有一定的优越性。

5 结束语

为了提高 QPSO 算法的性能,融入了变异机制和动态惯性权重形成 IQPSO。在 UCI 和课程成绩数据集上,将提出的 IQPSO 与 QPSO、Apriori 规则挖掘算法进行比较,证明 IQPSO 算法具有快速收敛的能力,且所挖掘的规则比 Apriori 算法更为合理。

未来会将提出的算法应用于更大规模的数据集中,通过对算法进行优化进一步提高执行效率。此外,通过结合其他的机器学习算法,进一步提高挖掘到的规则的价值。

参考文献:

[1]赵学健,孙知信,袁源.基于预判筛选的高效关联规则挖掘算法[J].电子与信息学报,2016,38(7):1654-1659.  
ZHAO Xuejian,SUN Zhixin,YUAN Yuan.An efficient association rule mining algorithm based on prejudging and screening [J].Journal of Electronics & Information Technology,2016,38(7):1654-1659.

[2]于守健,周昇阳.基于前缀项集的 Apriori 算法改进[J].计算机应用与软件,2017,34(2):290-294.  
YU Shoujian,ZHOU Yiyang.The improvement of Apriori algorithm based on prefixed-itemset[J].Computer Applications and Software,2017,34(2):290-294.

[3]HAHSLER M,KARPIENKO R.Visualizing association rules in hierarchical groups[J].Journal of Business Economics, 2017,87(3):317-335.

[4]刘军煜,贾修一.一种利用关联规则挖掘的多标记分类算法[J].软件学报,2017,28(11):2865-2878.  
LIU Junyu,JIA Xiuyi.Multi-label classification algorithm based on association rule mining[J].Journal of Software,2017,28 (11):2865-2878.

[5]WU X,WU S.An elitist quantum-inspired evolutionary algorithm for the flexible job-shop scheduling problem[J].Journal of Intelligent Manufacturing,2017,28(6):1441-1457.

[6]雷华军,秦开宇.测试不可靠条件下基于量子进化算法的测试优化选择[J].电子学报,2017,45(10):154-162.  
LEI Huajun,QIN Kaiyu.Optimal selection of imperfect tests based on improved quantum-inspired evolutionary algorithm [J].Acta Electronica Sinica,2017,45(10):154-162.

[7]TIRUMALA S S.A quantum-inspired evolutionary algorithm using Gaussian distribution-based quantization[J].Arabian Journal for Science & Engineering,2018,43(2):471-482.

表 5 前 10 条规则的平均适应度

Tab. 5 Average fitness of the top 10 rules

算法	幼儿园数据库	学生成绩数据集
Apriori <sup>[16]</sup>	0.376	0.387
GA <sup>[17]</sup>	0.387	0.395
IQPSO	0.403	0.416

- [8] MENG Y, LIU X. Quantum inspired evolutionary algorithm for community detection in complex networks[J]. Physics Letters A, 2018, 382(34): 2305-2312.
- [9] XI M, SUN J, LI L, et al. Cancer feature selection and classification using a binary quantum-behaved particle swarm optimization and support vector machine[J]. Computational & Mathematical Methods in Medicine, 2016, 16(9): 1-9.
- [10] ARAUJO D L A D, LOPES H S, FREITAS A A. Parallel genetic algorithm for rule discovery in large databases[C]// IEEE International Conference on Systems, 1999: 940-945.
- [11] LI J, WEI L, YING H. A new quantum rotation angle of quantum-inspired evolutionary algorithm for TSP[C]// IEEE International Conference on Computational Science & Engineering, 2017: 231-235.
- [12] SAMANTA S, CHOUDHURY A, DEY N, et al. Quantum-inspired evolutionary algorithm for scaling factor optimization during manifold medical information embedding[J]. Research and Applications, 2017: 285-326.
- [13] 刘小园. 量子粒子群算法在 WSN 三维定位中的研究[J]. 计算机应用与软件, 2018, 35(3): 120-123.  
LIU Xiaoyuan. Research on quantum particle swarm optimization algorithm in three dimensional localization of WSN[J]. Computer Applications and Software, 2018, 35(3): 120-123.
- [14] SUN J, LAI C H, XU W, et al. A novel and more efficient search strategy of quantum-behaved particle swarm optimization [C]// International Conference on Adaptive & Natural Computing Algorithms, 2007: 142-146.
- [15] 李浩君, 张广, 王万良. 一种惯性权重与种群多样性协同调整的二进制粒子群优化算法[J]. 小型微型计算机系统, 2018, 39(3): 529-533.  
LI Haojun, ZHANG Guang, WANG Wanliang. Binary particle swarm optimization algorithm for coordinated adjustment of population diversity and inertia weight[J]. Journal of Chinese Computer Systems, 2018, 39(3): 529-533.
- [16] ZHANG R, CHEN W, HSU T C, et al. ANG: A combination of Apriori and graph computing techniques for frequent itemsets mining[J]. Journal of Supercomputing, 2017, 32(3): 1-16.
- [17] KABIR M M J, XU S, KANG B H, et al. A new multiple seed based genetic algorithm for discovering a set of interesting Boolean association rules[J]. Expert Systems with Applications: An International Journal, 2017, 74(3): 55-69.

(责任编辑: 傅 游)