

基于二进制粒子群优化的卫星典型件工艺知识挖掘

王琳¹, 张永健², 钟诗胜¹, 刘金山³
(1. 哈尔滨工业大学机电工程学院, 黑龙江 哈尔滨 150001; 2. 哈尔滨工业大学(威海)船舶与海洋工程学院, 山东 威海 264209; 3. 中国空间技术研究院北京卫星制造厂, 北京 100094)

摘 要: 针对卫星典型件在工艺设计过程中设计任务量大、重复性工作多,且其历史工艺数据未能充分有效利用的问题,进行了工艺知识挖掘的研究,以提高工艺知识的重用性. 首先对工艺知识挖掘问题进行了描述,建立了工艺知识的关联规则模型;然后针对海量数据中 Apriori 算法挖掘效率低的问题引入二进制粒子群优化(BPSO)算法,并构造了基于 BPSO 的关联规则挖掘算法. 最后对卫星结构板这一典型件的历史工艺数据进行挖掘,得到了卫星结构板典型工序序列. 基于 BPSO 的关联规则挖掘算法可以有效提高工艺知识的挖掘效率.

关 键 词: 粒子群优化;典型件;工艺知识;关联规则挖掘;典型工序序列

中图分类号: TP 391 文献标志码: A 文章编号: 1005-3026(2015)01-0119-05

Binary Particle Swarm Optimization Based Process Knowledge Mining for Typical Parts of Satellite

WANG Lin¹, ZHANG Yong-jian², ZHONG Shi-sheng¹, LIU Jin-shan³
(1. School of Mechanical & Electrical Engineering, Harbin Institute of Technology, Harbin 150001, China; 2. School of Naval Architecture and Ocean Engineering, Harbin Institute of Technology at Weihai, Weihai 264209, China; 3. Beijing Satellite Manufacturing Factory, China Academy of Space Technology, Beijing 100094, China. Corresponding author: ZHANG Yong-jian, E-mail: mailoz@163.com)

Abstract: The huge quantity of design of manufacturing process of satellite typical parts and a lot of repeated jobs existed in the process. Many kinds of process knowledge without reused effectively contained in the historical process data. Process knowledge mining algorithm was studied in order to increase efficiency. The problem was described firstly, and the association rule model was built. In order to improve computational efficiency of Apriori algorithm for huge datasets, binary particle swarm optimization(BPSO) was introduced. Meanwhile association rule mining algorithm based on BPSO was designed. Finally, the designed algorithm was used in process knowledge mining for satellite plate. The mining efficiency of process knowledge can be improved effectively by the association rule mining algorithm based on BPSO.

Key words: particle swarm optimization; typical part; process knowledge; association rule mining; typical operation sequence

随着市场竞争的日益激烈,企业的响应速度成为在竞争中占据优势的关键,对此,各企业采用各种技术手段以提高自身的竞争力. 由于企业中存有大量的设计信息、制造信息等历史数据,丰富的产品知识蕴含于其中,对产品知识的挖掘与重用已成为企业所采用的重要手段之一^[1].

在众多的知识挖掘工具中,关联规则挖掘在产品的知识挖掘中常被采用,包括产品族设计知识的挖掘,产品设计与制造的映射关系识别,生产过程中质量控制规则的挖掘,以及工艺序列知识的挖掘^[2-5]等. Apriori 算法是关联规则挖掘的经典算法,但在挖掘过程中会产生大量的候选集,要

对数据库进行多次读取,当数据量较大时效率低.为提高效率,遗传算法、粒子群算法等方法被引入关联规则的挖掘算法中^[6-7].此外蚁群优化算法和 FP-growth 算法也是频繁模式挖掘的常用算法^[8-9],前者在数据量大时会出现迭代次数多、效率低的问题;FP-growth 算法只对数据库扫描两次,不产生大量的候选集,但该算法用于文本关联规则挖掘时,在时间和空间的效率不理想^[10].

卫星典型件是指在各型号卫星上都经常使用且结构功能相似的典型零部件,如结构板、管路等. 我国的卫星发射任务激增,典型件需求量巨大,当前对典型件的工艺设计仍采用传统的方式,工作量巨大且设计效率低. 由于卫星典型件结构相似,其工艺过程基本一致,为提高工艺知识的挖掘效率,本文将采用基于二进制粒子群的关联规则挖掘算法,从卫星典型件的历史工艺数据中抽取其工艺知识以支持卫星典型件的快速工艺设计.

1 工艺知识挖掘问题描述

在工艺数据库中存在大量的工艺对象,包括零件类型、材料、零件特征、加工方法、设备、刀具等,这些工艺对象之间存在着关联关系,如零件特征与加工方法、材料与刀具等. 这些关联关系中比较典型的或工艺设计人员经常采用的都可作为工艺知识. 这些工艺知识可以根据其必要性和适应性分为硬知识和软知识,其中硬知识是显而易见的、必要的或强制性的,而软知识则与工艺人员的经验、企业制造能力、客户要求等相关^[6]. 因此,软知识是工艺知识挖掘的主要目标.

1.1 工艺知识的关联规则模型

对工艺知识挖掘而言,给定一事务集 $T = \{t_1, t_2, \dots, t_n\}$,该事务集是某类型零部件的工艺集合, t_i 是每个事务的数据, X 称为数据项,每个工艺过程(工序工步)都是一个数据项. 事务集中的所有数据构成数据项集 $I = \{i_1, i_2, \dots, i_k\}$. 关联规则是形如 $X \Rightarrow Y$ 的逻辑蕴含式,其中 $X \subseteq I$, $Y \subseteq I$,且 $X \cap Y = \emptyset$, X 称为前项, Y 称为后项,如机加工工艺规程中“先粗后精”的加工顺序即可作为一条工艺知识规则“粗加工 \Rightarrow 精加工”.

项集 X 和项集 Y 中所包含的项在事务集 T 中同时出现的概率记为关联规则的支持度 $\text{Sup}(X \Rightarrow Y)$,是对关联规则重要性(或适用范围)的衡量.

$$\text{Sup}(X \Rightarrow Y) = \text{Sup}(X \cup Y) = P(X \cup Y).$$

在包含项集 X 的事务中,出现项集 Y 的条件概率记为关联规则的置信度 $\text{Conf}(X \Rightarrow Y)$,是对关联规则准确度的衡量,用以度量关联规则的强度.

$$\text{Conf}(X \Rightarrow Y) = \text{Sup}(X \cup Y) / \text{Sup}(X) = P(Y|X).$$

设定最小支持度 min-Sup 和最小置信度 min-Conf 两个阈值,若项集的支持度大于或等于最小支持度 min-Sup ,则称为频繁项集;若频繁项集的置信度大于或等于最小置信度 min-Conf ,则该频繁项集称为强规则. 因此,挖掘事务数据库中关联规则的问题可以划分为以下两个过程:①找出所有满足最小支持度的项集,即获得频繁项集;②通过频繁项集产生强关联规则.

关联规则挖掘的经典算法是 Apriori 算法,该算法要对数据库作多次遍历,对小规模的数据库该算法完全能够满足,而对数据量大的数据库该算法效率低、耗时多. 对此,引入粒子群优化算法以提高大数据量数据库的关联规则挖掘的效率^[8].

1.2 二进制粒子群优化算法

粒子群优化 (particle swarm optimization, PSO) 算法是由 Kennedy 等于 1995 年提出的一种群体智能演化算法,最初用于模拟鸟群觅食的过程,如图 1 所示,后来经过不断地改进成为一种很好的优化工具,在多目标优化、模式识别、决策支持、数据挖掘等领域得到广泛应用.

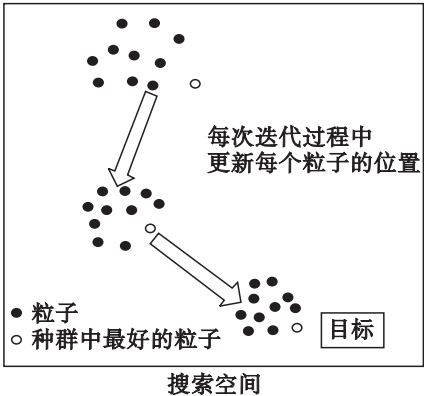


图 1 种群跟随最优的粒子移动到目标
Fig. 1 Swarm following the best particle to move to the goal

PSO 算法包括 3 个步骤:生成粒子的初始位置和速度、粒子速度更新、粒子位置更新. 在 PSO 算法中,每个被称为粒子的个体或优化问题的可能解在搜索空间中都有一个运动速度,该速度根据这个粒子自身的经验和相邻粒子的经验或者群体的经验来动态调整,粒子 i 的速度迭代公式和位置迭代公式为

$$V_{i,k+1} = W \cdot V_{i,k} + c_1 r_1 (P_{i,k} - X_{i,k}) + c_2 r_2 (G_k - X_{i,k}),$$
$$X_{i,k+1} = X_{i,k} + V_{i,k}.$$

其中： $V_{i,k+1}$ 为粒子*i*更新后的速度； $X_{i,k+1}$ 为粒子*i*更新后的位置； $X_{i,k}$ 为粒子*i*的当前位置； $P_{i,k}$ 为粒子*i*的个体最优位置； G_k 为整个种群的全局最优位置； W 为惯性权重，调整全局搜索能力与局部搜索能力之间的变化，通常取值范围为 $[0,1]$ ； c_1 ， c_2 为学习因子； r_1 ， r_2 为 $(0,1)$ 之间的随机数。

PSO 算法最初用于解决连续空间内的优化问题，而实际的工程应用问题多是离散的或组合优化问题，为此二进制粒子群优化(binary particle swarm optimization, BPSO)算法被提出并应用于实际的工程问题中^[7]。BPSO 算法中粒子速度迭代公式与实际 PSO 算法中的速度迭代公式形式上是相同的，但意义完全不同。粒子位置迭代公式如下：

$$X_{i,k+1} = \begin{cases} 1, & r_3 < \text{sig}(V_{i,k+1}); \\ 0, & r_3 \geq \text{sig}(V_{i,k+1}). \end{cases}$$

其中： r_3 为 $(0,1)$ 之间的随机数； $\text{sig}(\cdot)$ 为 Sigmoid 函数，通常取 $\text{sig}(x) = 1/(1 + \exp(-x))$ 。

2 基于 BPSO 的关联规则挖掘算法

基于 BPSO 的关联规则挖掘算法包括两个部分：数据准备和规则挖掘。数据准备部分主要是计算粒子群的适应度，将数据库中的数据转化为二进制格式并保存；然后利用 PSO 算法进行关联规则挖掘。首先对粒子群进行编码，然后根据适应度生成特定群体，最后启动搜索程序直到满足约束条件，即找到最优粒子。该粒子的支持度和置信度即为最小支持度和最小置信度，用于后续的关联规则挖掘。

2.1 二进制转换

首先将事务集中的数据转换成二进制格式，即每条数据的记录和存储都用“0”或“1”表示。这种方法可以加快数据库的扫描速度，而且支持度和置信度的计算也更加简单方便。这种变换方法可通过一个例子进行说明，如图 2 所示。在原始数据库中有从 T_1 到 T_5 共 5 条记录，每条记录都可以转换为二进制形式，并以二进制形式进行存储。例如，在数据库中一共有 5 种不同的工艺对象，每条工艺路线都有 5 个单元，以 B_2 为例，在这条工艺路线中有 3 个工艺对象分别为 I_1 ， I_2 和 I_3 ，因此与之对应的单元中都是“1”，而其他单元为“0”。

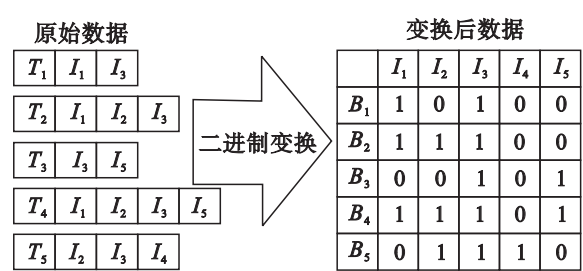


图 2 数据的二进制转换
Fig. 2 Binary transformation of data

2.2 关联规则编码

由于遗传算法在规则产生领域有成功的应用，参照其中规则的编码方法，针对工艺知识的特点，采用密西根方法(Michigan approach)对关联规则进行编码。

假如数据库中存在 N 个项目，每个项目包括两个部分，每个部分有两个可能值“0”或“1”。对于第一部分而言“1”表示该项目在某条规则中出现，“0”则表示该项目没有在规则中出现。第二部分的值表示该项目是否属于某规则的前项或后项，如果值为“1”则表示该项目出现在某规则的前项中，如果值为“0”则表示该项目出现在某规则的后项中。由此可见，每个粒子的维数为 $2N$ ，在 BPSO 算法中关联规则的编码如表 1 所示。

表 1 关联规则在 BPSO 算法中的编码						
Table 1 Association rule encoding in BPSO algorithm						
I_1		I_2		...	I_N	
V_{11}	V_{12}	V_{21}	V_{22}	...	V_{N1}	V_{N2}

2.3 适应度计算

在 PSO 算法中，适应度这个概念是用来评价每个粒子在优化计算中可能达到或有助于找到最优解的重要性，适应度越高则粒子的位置越好，因此，适应度最大的种群可作为原始种群。每个粒子的适应度来源于适应度函数，适应度函数必须有效地反映每个粒子与最优解之间的距离，适应度函数的确定直接影响 PSO 算法的搜索性能与计算效率，根据工艺数据库中工艺规则的特点，本文中关联规则 $X \Rightarrow Y$ 的适应度函数表示为

$$\text{Fitness} = a \cdot \text{Sup}(X \Rightarrow Y) + b \cdot \text{Conf}(X \Rightarrow Y).$$

其中： $\text{Sup}(X \Rightarrow Y)$ 为关联规则的支持度； $\text{Conf}(X \Rightarrow Y)$ 为关联规则的置信度； a ， b 为常数，取值范围为 $[0,1]$ ，用于调整关联规则评价的偏重。

2.4 关联规则挖掘流程

基于 BPSO 的关联规则挖掘算法基本流程如图 3 所示。该过程包含两部分：数据预处理和规则挖掘。数据预处理主要涉及数据二进制转换、适应度计算

等步骤;规则挖掘部分首先通过规则编码和粒子适应度确定原始种群,然后应用 PSO 算法进行搜索,获得粒子的全局最优位置,该位置粒子的支持度和置信度即可作为关联规则挖掘的最小支持度与最小置信度.同时给出 PSO 算法的基本实现过程.

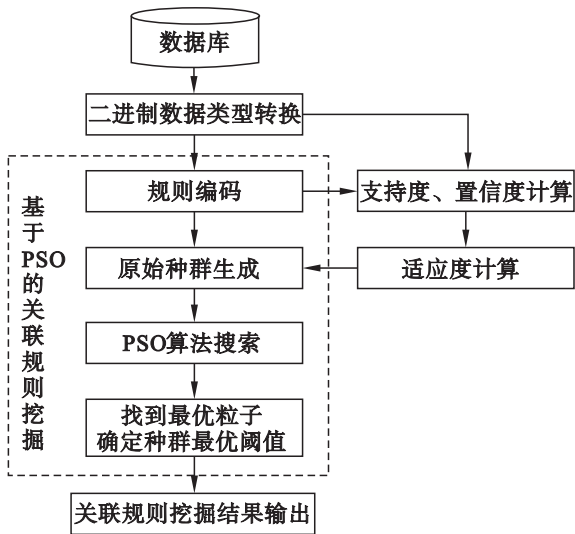


图 3 基于 BPSO 的关联规则挖掘算法流程
Fig. 3 BPSO based association rule mining process

PSO 算法:
输入: n (粒子规模), m (粒子维数), c_1, c_2 (学习因子), w (权重)
输出:PGbest(全局最优位置)
//初始化粒子的位置和速度
For $i = 1$ to n
 For $j = 1$ to m
 $X[i][j] = Llimit + (Ulimit - Llimit) \times r$
 $V[i][j] = 0$
 End
End
//初始化全局和局部适应度
 $fitness_Gbest = inf$
For $i = 1$ to n
 $fitness_Lbest[i] = inf$
End
 $Fitness_X = evaluate_fitness(X)$
//更新局部最优位置和适应度
For $i = 1$ to n
 If ($fitness_X(i) < fitness_Lbest(i)$)
 $fitness_Lbest[i] = fitness_X(i)$
 For $j = 1$ to m
 $X_Lbest[i, j] = X[i, j]$
 End
End

End
//更新全局最优位置和适应度
 $[min_fitness, min_fitness_index] = min(fitness_X(i))$
If ($min_fitness < fitness_Gbest$) $fitness_Gbest = min_fitness$
 For $j = 1$ to m
 $X_Gbest[j] = X(min_fitness_index, j)$
 End
End
//更新粒子的位置和速度
For $i = 1$ to n
 For $j = 1$ to m
 $V[i][j] = w \times V[i][j] + c_1 \times r_1 \times (X_Lbest[i][j] - X[i][j]) + c_2 \times r_2 \times (X_Gbest[j] - X[i][j])$
 $X[i][j] = X[i][j] + V[i][j]$
 End
End

3 卫星典型件工艺知识挖掘实例

以卫星结构板为例,在其工艺编制过程中,很多工艺单元是重复使用的,将这些成熟的、相对固定的若干工序(工步)组成标准的工艺规程单元,即工艺知识中的硬知识,可直接应用于工艺编制中.根据结构板材料的不同,其研制流程涉及 6 ~ 7 种工艺,典型工序序列是多种工艺知识中有代表性的,因此,本文以复合材料结构板典型工艺序列的挖掘为算例.对某卫星制造企业 CAPP 数据库中的复合材料结构板成型工艺进行整理,形成如表 2 所示的工艺路线数据集,为简化计算过程,只取 10 个工艺路线事务,直接用工序名称作为数据项,共 135 个数据项.从中可以看出,诸如“准备→备料”、“车铣→阳化”等工序序列可直接作为典型工序序列.

用前文研究的算法对表 2 中的数据进行典型工序序列的挖掘,通过计算得到全局最优位置处粒子的支持度 Sup 为 0.58,置信度为 0.62,取表 3 中的三组支持度和置信度分别用 Apriori 算法和 BPSO 算法进行规则挖掘,结果如表 3 所示. Apriori 算法挖掘得到的规则数目多于 BPSO 算法,是因为 BPSO 算法是随机搜索算法,会导致规则挖掘不全,但可以得到大部分的规则,而 Apriori 算法与 BPSO 算法相比耗时多,因此 BPSO 算法在提高规则的挖掘效率方面优势明

显. 受限于篇幅,不再列出具体结果,之所以会出现如此多的强规则是因为多个工序不同的排列组合造成的,对工艺本身而言多数规则实质上是相

同的,而且对典型件的工艺也是有效的. 此外粒子群算法与遗传算法在关联规则挖掘效率方面的比较可参考文献[11],本文不再赘述.

表 2 结构板成型工艺路线数据集
Table 2 Dataset of shaping process for satellite plate

序号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
P ₁	准备	备料	试件	工装	试装	车铣	阳化	装配	固化	修整	理化	后装	胶钳	检	入库
P ₂	准备	备料	车铣	阳化	试装	车铣	阳化	装配	固化	修整	理化	胶钳	检	入库	—
P ₃	准备	备料	试件	工装	试装	装配	封装	固化	修整	理化	胶钳	检	入库	—	—
P ₄	准备	备料	车铣	阳化	试件	试装	工装	装配	固化	修整	理化	胶钳	检	入库	—
P ₅	准备	备料	试件	试装	车铣	阳化	工装	装配	固化	修整	理化	后装	胶钳	检	入库
P ₆	准备	备料	试件	工装	试装	装配	固化	修整	理化	胶钳	检	入库	—	—	—
P ₇	准备	备料	试件	车铣	阳化	工装	装配	封装	固化	修整	理化	胶钳	检	入库	—
P ₈	准备	备料	试件	试装	工装	装配	固化	修整	理化	后装	胶钳	检	入库	—	—
P ₉	准备	备料	工装	装配	封装	固化	修整	理化	后装	胶钳	检	入库	—	—	—
P ₁₀	准备	备料	工装	试装	车铣	阳化	装配	固化	修整	理化	胶钳	检	入库	—	—

表 3 BPSO 算法与 Apriori 算法挖掘结果比较

Table 3 Comparison of mining result between Apriori algorithm and BPSO algorithm

挖掘算法	支持度	置信度	规则数	耗时/s
Apriori 算法	0. 6	0. 6	4 094	22 010. 3
	0. 8	0. 8	1 535	606. 7
	1	0. 8	510	78. 2
BPSO 算法	0. 6	0. 6	3 824	13 097. 2
	0. 8	0. 8	1 302	438. 5
	1	0. 8	479	65. 7

4 结 论

- 1) 关联规则挖掘算法可用于工艺知识的挖掘,而且工艺知识能较好地用关联规则进行表达.
- 2) 基于 BPSO 的关联规则挖掘算法可以有效提高工艺知识的挖掘效率.
- 3) 本文只用典型工序序列的挖掘对挖掘算法进行了验证,实际上工艺历史数据中所蕴含的知识更丰富;由于工艺历史数据的不规范性,为方便挖掘,在进行知识挖掘之前需要对其进行整理,此过程同样值得研究.

参考文献:

[1] Gülser K, İnci B, Murat T. A review of data mining applications for quality improvement in manufacturing industry[J]. *Expert Systems with Applications*,2011,38(1): 13448 – 13467.

[2] Moon S K, Simpson T W, Kumara S R T. A methodology for knowledge discovery to support product family design[J]. *Annals of Operations Research*,2010,174(1):201 – 218.

[3] Zhang L L, Jiao R J. Identifying mapping relationships

between functions and technologies: an approach based on association rule mining [C]// International Conference on Industrial Engineering and Engineering Management (IEEM). Singapore:IEEE,2011:1596 – 1601.

[4] Kamsu-Foguem B,Rigal F,Mauget F. Mining association rules for the quality improvement of the production process [J]. *Expert Systems with Applications*,2013,40(4):1034 – 1045.

[5] 高伟,殷国富,成尔京. 机械制造工艺序列中的知识发现方法研究[J]. *机械工程学报*,2004,40(5):121 – 125,130. (Gao Wei, Yin Guo-fu, Cheng Er-jing. Research on method of knowledge discovery in manufacturing process sequence [J]. *Chinese Journal of Mechanical Engineering*, 2004, 40 (5):121 – 125,130.)

[6] Li C R, Yang C X, Xiong P. Implementation of the multi-agent-based product knowledge extraction system[C]// WRI Global Congress on Intelligent Systems. Xiamen,2009:548 – 552.

[7] Sarath K N V D, Ravi V. Association rule mining using binary particle swarm optimization [J]. *Engineering Applications of Artificial Intelligence*, 2013, 26 (8): 1832 – 1840.

[8] Patel B, Chaudhary V K, Karan R K, et al. Optimization of association rule mining Apriori algorithm using ACO [J]. *International Journal of Soft Computing and Engineering*, 2011, 1 (1): 24 – 26.

[9] Soni R K, Gupta N, Sinhal A. An FP-growth approach to mining association rules [J]. *International Journal of Computer Science and Mobile Computing*, 2013, 2 (2): 1 – 5.

[10] 黄嘉满,张冬荣. 基于文本的关联规则提取方法的研究 [J]. *计算机仿真*,2008,25(1):96 – 99. (Huang Jia-man, Zhang Dong-mo. An algorithm for mining association rule in textual database[J]. *Computer Simulation*, 2008, 25 (1): 96 – 99.)

[11] Indira K, Kanmani S, Prashanth P, et al. Population based search methods in mining association rules [C]// The Third International Conference, CNC 2012. Chennai, 2012:255 – 261.