

关联规则在提取典型焊接类型中的应用^①

晁永生^② 刘海江

(同济大学机械工程学院 上海 201804)

摘 要 为了从白车身工艺规划数据库中提取潜在的、有价值的焊接工艺知识,应用关联规则提取了典型焊接类型。首先建立了提取典型焊接类型的关联规则模型,然后对影响焊接类型的各种属性分类,并根据焊接工艺要求对数值属性离散化,采用 Apriori 算法提取频繁属性项集。将分属不同属性的项连接生成候选项集,以焊接类型作为规则的后件,根据最小置信度获取强规则。最后对某白车身的焊接数据进行处理,为了能够提取合适的焊接类型,采用较低的支持度和较高的置信度,获取的结果对焊接类型的选择有很强的参考价值。这种方法有助于实现白车身工艺规划的智能化,提高了工艺规划的效率。

关键词 关联规则, Apriori 算法, 焊接类型, 知识发现, 白车身, 工艺规划

0 引 言

计算机辅助工艺设计 (computer aided process planning, CAPP) 需要从积累的大量工艺设计数据中获取有用的工艺知识,但目前对这些知识的获取以及将知识表达成易于 CAPP 系统理解的形式,仍是 CAPP 系统开发的主要瓶颈之一^[1]。采用人工方式将会花费大量人力、物力、财力和时间,而且难以进行知识更新和扩充,严重影响知识的重用^[2]。近年来数据挖掘^[3]已被广泛应用于 CAPP 中,用以实现自动获取工艺知识,刘书暖^[4]等人提出了基于聚类分析的典型工艺路线发现方法;刘闯^[5]等人运用灰度关联方法来检索钣金 CAPP 中相似的工艺实例;高伟^[6]等人提出了基于序列模式的机械制造工艺序列中的知识发现方法。这些方法主要是发现工艺数据库中的典型工艺序列,并没有涉及具体的工艺决策。在工艺规划中,决定使用某一工艺的因素很多,在不同的条件下使用不同的工艺,影响工艺的各个因素之间存在一定的关联。本文将关联规则的挖掘思想应用于白车身工艺规划中,以获取典型的焊接类型。

1 关联规则用于白车身工艺规划中典型焊接类型获取的思想

轿车车身通常由 300 ~ 500 个具有复杂空间曲

面的薄板冲压零件组成。在白车身制造工艺中,焊接是将各个零件连接在一起的最常用的方法。汽车的轻量化是当代汽车创新工程的技术发展方向,各种轻量化的合金材料已被广泛使用在车身各个部位。不同的材料需要采用不同的焊接类型,同时不同的工艺要求也需要采用相应的焊接类型。焊接类型的选择直接关系到两个部件是否能够连接以及连接后变形的大小,能否满足工艺性要求,同时焊接类型还影响焊接的加工成本。特别是合金材料相对传统的钢材对焊接类型要求更加严格,焊接类型对其焊接质量影响更大。因此,焊接类型选择在白车身工艺规划中非常重要。

一个白车身的焊点达 4000 ~ 5000 个,在工艺规划中涉及大量的焊接信息,同时在已完成的工艺规划中,也留下了大量的焊接工艺数据,这些数据是以往工艺经验的总结,是进行新的工艺设计的宝贵财富。从中发现各种工艺信息之间的关系,对进行新的设计可起指导作用。而关联规则挖掘是目前数据挖掘领域中最为活跃的研究方向之一,它是用来分析数据库中不同属性数据间有价值的相互依存性和关联性。关联是指两个或多个变量的取值之间存在某种规律性。如果两项或者多项属性之间存在一定的关联关系,那么其中的一项属性值就可以依据其他属性值进行预测。影响焊接类型选择的因素很多,包括焊接的材料、材料的厚度、焊缝的密封性要

^① 863 计划(2008AA04Z105)资助项目。

^② 男,1976 年生,博士;研究方向:数字化制造技术,汽车车身工艺规划;联系人, E-mail: cys21st@yahoo.com.cn
(收稿日期:2009-04-27)

求等。每一个影响因素可以看作一个属性,可以将这些属性连同焊接类型看作一个焊接属性项集,运用关联规则发现其中可信度较高的约束关联。如在焊接车门外板时,如果使用的材料是普通的中碳钢,厚度是0.8mm,没有密封性的要求,通常会选择经济性较好的点焊,这就是各个属性之间的一种紧密的联系。所以典型焊接类型获取实质上就是发现与焊接类型属性关联程度较高的属性项集。

2 典型焊接类型获取的关联规则和Apriori算法

2.1 关联规则

关联规则挖掘技术用于发现数据库中属性之间有趣、潜在、隐含的联系^[7]。设有项目集 $I = \{i_1, i_2, \dots, i_m\}$ 是所有数据项的集合, D 是事务数据库,其中每一个事务 T 是 I 中一些项目的集合,即 $T \subseteq I$,对于项目集 $X \subseteq I$,若 $X \subseteq T$,则事务 T 支持 X 。

定义1 在数据库 D 中,如果事务 T 包含数据项集 X 的所有子集,称 T 支持 X ,数据项集的支持度 $Supp(X)$ 是指 D 中支持 X 的事务和所有事务数量的比值。

定义2 设 X, Y 为数据库 D 中不相交的数据项集, $X, Y \subseteq I, X \cap Y = \emptyset$,则关联规则是如下形式的一个蕴含式: $R: X \rightarrow Y$ 。

定义3 规则 R 的置信度为 c ,是指在数据库 D 中支持 X 的事务中有 $c\%$ 的事务也支持 Y ,其基于条件概率的定义为 $Conf(R) = \frac{p(Y \subseteq T \wedge X \subseteq T)}{p(X \subseteq T)}$

$$= \frac{Supp(X \cup Y)}{Supp(X)}。$$

定义4 规则 R 的支持度为 s ,是指在数据库 D 中有 $s\%$ 的事务包含 $X \cup Y$ 。

定义5 如果给出最小支持度阈值 $minsup$ 和最小置信度 $minconf$,若 $Supp(R) \geq minsup$ 且 $Conf(R) \geq minconf$,则称 R 为强规则。

定义6 对于一个给定的最小支持度 $minsup$,如果数据项集 X 的支持度 $Supp(X) > minsup$,则称 X 是频繁的,并称其为频繁数据项集(frequent item-set)。

与白车身焊接工艺相关的数据种类很多,其属性值可分为布尔型和多值型。多值型属性又可分为类别属性和数量属性,类别属性包括焊接类型、材料的种类等,每个属性值就是项目集 I 中的一项。数值属性包括如零件的厚度、尺寸等。数值属性在挖

掘之前必须对数据离散化,将数值属性的值用相应的区间来代替。对于连续的属性 A_i ,若属性空间 $[u_i, v_i]$ 被分为 n 个子区间 $[u_{i1}, v_{i1}], [u_{i2}, v_{i2}], \dots, [u_{in}, v_{in}]$,则属性 A_i 将拆分为 n 个项 $\langle A_i, u_{i1}, v_{i1} \rangle, \langle A_i, u_{i2}, v_{i2} \rangle, \dots, \langle A_i, u_{in}, v_{in} \rangle$ 。所有布尔属性值、类别属性值和数值属性值共同构成项目集 I 。每个零件所有的属性值就构成了一个事务 T 。

关联规则挖掘的任务是找出 D 中所有的强规则。可把关联规则挖掘划分为以下两个子问题:(1)根据最小支持度找出数据集 D 中所有的频繁项集;(2)根据频繁项目集和最小置信度产生关联规则。典型焊接类型提取关联规则挖掘的基本模型如图1。

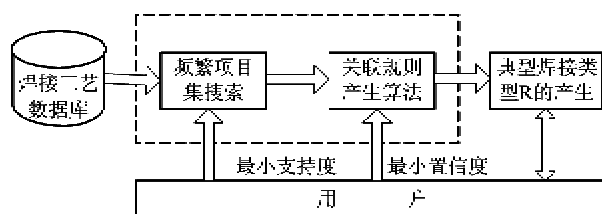


图1 关联规则挖掘的基本模型

2.2 典型焊接类型获取的Apriori算法

目前关联规则的新算法、新应用层出不穷,相关问题定义和背景也不尽相同。在众多算法中,由Agrawal和Srikant提出的Apriori算法是最著名算法之一^[7]。因此,本次白车身工艺规划中的典型焊接发现就采用Apriori算法。Apriori算法需对数据库做多次遍历,其核心是基于频繁项集理论的递推方法,即 k 项集用于搜索 $(k+1)$ 项集。基于关联规则的挖掘只对在数据库记录中频繁出现的数据项集感兴趣。数据项集有以下几个非常有用的性质:(1)频繁项集的所有非空子集都必须是频繁的;(2)非频繁项集的超集一定是非频繁的;(3)对于数据项集 A 和 B ,如果 $A \subseteq B$,则 $Supp(A) \geq Supp(B)$ 。这几个性质是关联算法的设计基础。

为了压缩搜索空间,用Apriori算法提取典型焊接类型的具体实现过程如下:

(1)生成频繁项集 L_1 。通过单趟扫描数据事务数据库 D ,计算出各1项集的支持度,计算出各个项的支持度,即扫描焊接数据集,计算焊接件各属性值的支持度。大于最小支持度项就构成了频繁1项集 L_1 。

(2)联接过程。为了产生频繁 k 项集 L_k ,需要

预先生成一个候选 k 项集 C_k 。在 $k-1$ 频繁数据项集 L_{k-1} 中搜索所有前 $k-2$ 数据项相同、最后一项不同的数据项集对相互联接生成一个新的 k 项候选数据项。为了避免重复,联接的第二项数据项集的最后一个数据项应大于第一个数据项的相应数据项。由 L_{k-1} 联接生成的项集就是 C_k 。

(3)剪枝过程。由于 C_k 是 L_k 的超集,可能有些项不是频繁的。 C_k 通常很庞大,由 C_k 生成 L_k 会带来巨大的计算量,为减少 C_k 的规模,Apriori 遵从性质“任何非频繁的 $(k-1)$ 项集必定不是频繁 k 项集的子集”。所以,当在 k 项集的某个 $(k-1)$ 子集不是 L_{k-1} 的成员时,则该候选项集不可能是频繁的,可以从 C_k 中删去。

(4)计算 C_k 中各个项集的支持度。通过遍历扫描数据事务数据库 D ,计算出各个 C_k 中各个项集的支持度,即通过扫描焊接数据记录集,计算出每个项集的支持度。为了减少计算候选项集支持度所涉及的记录总数,在每次计算 C_k 支持度的过程中,给不包含在 C_k 中的任何项集的事务打上删除标记,在以后数据事务集 D 的扫描计数中不予考虑^[2]。

(5)生成频繁项集。将 C_k 中不满足最小支持度的项剔除,形成由频繁 k 项集构成的集合 L_k 。通过循环迭代,重复(1)-(4),直到不能产生频繁项集的集合(非空集合)为止,Apriori 算法求出所有满足最小支持度的频繁项集。

(6)产生规则。关联规则是在频繁项集中产生的。对于每一个频繁数据项集 l ,找出其所有的非空子集,对于每一个这样的非空子集 a ,如果 $Supp(l)$ 与 $Supp(a)$ 的比值大于最小支持度,就可以形成如下一条规则 $(l-a) \rightarrow a$ 。也就是说,将任意频繁数据项集的一个子集作为规则的后件(结论),其剩余部分作为规则前件(前提)。由频繁项集的任意子集都是频繁项集,根据定义 4 计算该规则的支持度,再利用定义 5 判断该规则是否为强规则。将焊接类型属性值作为规则后件,其余属性值作为规则前件,若是强规则,规则后件 a 就是获取的典型焊接类型。

3 典型焊接类型的获取实例

经过挖掘准备,本文以一轿车的一组零部件焊接工艺为例,挖掘典型焊接类型与其它焊接相关性之间的联系,来说明典型焊接类型的获取的过程。有 10 个焊接零件,即数据库中有 10 个事务,如表 1

所示。

表 1 10 个零件的焊接类型及其相关属性值

零件名称	焊接类型	焊接材质	密封性	经济性	厚度 (mm)
左前门外板	点焊	普通中碳钢	否	好	1.3
车棚顶盖	缝焊	普通中碳钢	是	好	0.8
横梁	CO ₂ 保护焊	普通中碳钢	否	好	1.8
车身底板	激光焊	铝合金	是	差	1.5
左前门内板	激光焊	高强度钢	否	差	1.0
发动机外罩	点焊	铝合金	否	好	1.0
后面板	点焊	普通中碳钢	否	好	0.8
地板通道	点焊	普通中碳钢	否	好	0.8
门窗框	激光焊	铝合金	是	差	1.0
行李舱	点焊	普通中碳钢	否	好	0.8

焊接类型很多,某些焊接类型使用次数相对较少,但可能是对某些具体的属性最合适。因此,为了将这些焊接类型提取,采取相对较低的支持度和高的置信度。这里取最小事务支持计数为 2,最小置信度为 70%。具体说明如下:

(1)在表 1 中与焊接有关的共有 5 个属性,焊接类型属于多值属性,共有 4 项,包括点焊、缝焊、CO₂ 保护焊和激光焊。焊接材料也是多值属性,共 3 项,密封性属布尔属性,共 2 项,经济性属于布尔属性,共 2 项。厚度属性是数字量,不同的焊接厚度可能采取不同的焊接类型,将厚度分为两项,即厚度(0.6~1.2)和厚度(1.2~3.0)。属性项集 I 共包括 13 项。

(2)在第一次扫描事务数据库 D 时,每一项都是候选 1 项集 C_1 的成员,对每一个项集进行了编号,计算各项集的支持数,如图 2(a)所示。根据最小支持数的要求,删除非频繁项集,获得频繁项集 L_1 ,如图 2(b)所示。

(3)来自同一属性的两项生成 2 项集也没有实际意义,其支持度也为 0,如项集 $\{it_1, it_9\}$ 。因此,从 L_1 中选择分属不同属性的项进行连接,这样可大大减少生成候选项集的数量。项集 L_1 连接后生成候选项集 C_2 ,计算各项集的支持数,删除非频繁项集,获得频繁项集 L_2 ,如图 2(c)所示。

(4)生成 3 项集时,首先考虑从 L_2 中选择首项相同的项集,末项不同,如 $\{it_1, it_2\}$ 和 $\{it_1, it_4\}$,可以生成 3 项集 $\{it_1, it_2, it_4\}$ 。不断遍历事务数据库 D ,依次生成频繁项集 L_3, L_4, L_5 ,分布如图 2(d), (e), (f)所示。生成 L_5 ,已经包含了所有相关的属性项,这时 $C_6 = \emptyset$,算法终止。

(5)产生规则。这里的规则后件是焊接类型,规则前件是其余4个属性。对于频繁项集 L_3 ,以点焊为规则后件的关联规则的置信度计算如下: {普通中碳钢,密封性(否)} $\rightarrow$ {点焊}, $Conf(R) = 80\%$; {普通中碳钢,经济性(好)} $\rightarrow$ {点焊}, $Conf(R) = 66.7\%$; {普通中碳钢,厚度(0.6~1.2)} $\rightarrow$ {点焊}, $Conf(R) = 75\%$; {密封性(否),经济性(好)} $\rightarrow$ {点

焊}, $Conf(R) = 83.3\%$; {密封性(否),厚度(0.6~1.2)} $\rightarrow$ {点焊}, $Conf(R) = 80\%$; {经济性(好),厚度(0.6~1.2)} $\rightarrow$ {点焊}, $Conf(R) = 80\%$ 。结合图 2(d),从上述置信度分析可以看出,除了规则 {普通中碳钢,经济性(好)} $\rightarrow$ {点焊},其余规则均满足置信度的要求,因此为强规则。

编号	候选项集	支持数
it_1	点焊	5
it_2	普通中碳钢	6
it_3	密封性(否)	3
it_4	密封性(是)	7
it_5	经济性(好)	7
it_6	经济性(差)	3
it_7	CO ₂ 保护焊	1
it_8	高强度钢	1
it_9	激光焊	3
it_{10}	缝焊	1
it_{11}	铝合金	3
it_{12}	厚度(0.6~1.2)	7
it_{13}	厚度(1.2~3)	3

(a) 候选项集 C_1

频繁项集	支持数
it_1	5
it_2	6
it_3	3
it_4	7
it_5	7
it_6	3
it_9	3
it_{11}	3
it_{12}	7
it_{13}	3

(b) 频繁项集 L_1

频繁项集	支持数
it_1, it_2	4
it_1, it_4	5
it_1, it_5	5
it_2, it_{12}	4
it_2, it_4	5
it_2, it_5	6
it_3, it_{12}	4
it_4, it_5	6
it_4, it_{12}	5
it_5, it_{11}	2
it_9, it_{11}	2
it_9, it_5	5
it_5, it_{12}	2
it_3, it_{11}	2
it_5, it_{12}	5
it_5, it_{12}	2

(c) 频繁项集 L_2

频繁项集	支持数
it_1, it_2, it_4	4
it_1, it_2, it_5	4
it_2, it_4, it_{12}	3
it_1, it_4, it_5	5
it_2, it_4, it_{12}	4
it_2, it_5, it_{12}	4
it_2, it_4, it_5	5
it_2, it_4, it_{12}	3
it_4, it_5, it_{12}	4
it_3, it_9, it_{11}	2
it_9, it_5, it_{11}	2
it_2, it_5, it_9	2
it_3, it_9, it_{11}	2

(d) 频繁项集 L_3

频繁项集	支持数
it_1, it_2, it_4, it_5	4
$it_1, it_2, it_4, it_{12}$	3
$it_2, it_4, it_5, it_{12}$	3
$it_2, it_5, it_9, it_{11}$	2

(e) 频繁项集 L_4

频繁项集	支持数
$it_1, it_2, it_4, it_5, it_{12}$	3

(f) 频繁项集 L_5

图 2 典型焊接类型获取的具体过程

对于频繁项集 L_4 和 L_5 ,关联规则置信度计算如下: {普通中碳钢,密封性(否),经济性(好)} $\rightarrow$ {点焊}, $Conf(R) = 80\%$; {普通中碳钢,密封性(否),厚度(0.6~1.2)} $\rightarrow$ {点焊}, $Conf(R) = 100\%$; {铝合金,密封性(是),经济性(差)} $\rightarrow$ {激光焊}, $Conf(R) = 100\%$; {普通中碳钢,密封性(否),经济性(好),厚度(0.6~1.2)} $\rightarrow$ {点焊}, $Conf(R) = 100\%$ 。

结合图 2(e)、图 2(f),从上述置信度分析可以

看出,上述规则均满足置信度的要求,因此均为强规则。

将上述所获取的强规则存放到知识库。采用产生式规则把得到的知识表示成“模式-动作”形式,如 $if(p_1, p_2, \dots, p_m) then(q)$,其中 $p_1, p_2, \dots, p_m$ 为规则的条件属性; q (即焊接类型)为规则中的决策属性。例如频繁项集 L_4 中的第一项可以表示为: if 材料 = “普通中碳钢” and 密封性 = “否” and 经济性 = “好” $then$ 焊接类型 = “点焊”,支持度 = 40%,置

信度 = 80%。当输入零件的材料、材料厚度、密封性和经济性要求时,就会从知识库中寻找相应的规则,提示可能的焊接类型,并给出相应规则的支持度和置信度。当不知道所有的属性值时,也可以只输入其中的几个值,系统会从对应的频繁项集(如 L_3 , L_4)中查找规则。

4 结论

计算机技术和通信技术的迅猛发展,使得各制造企业都在积极实行信息化,并且已经积累了大量的工艺数据。从中提取有用的信息并迅速运用到工艺规划中是对工艺规划人员的一个巨大挑战。因此,本文针对白车身工艺规划中焊接类型的选择与零件的材质、尺寸和工艺要求有密切的联系,运用关联规则和 Apriori 算法来获取的典型焊接类型,建立了典型焊接类型获取的关联规则模型,将分属不同属性的项进行连接生成候选项集,大大减少候选项数量,提高了 Apriori 算法的效率。最后给出了应用关联规则模型和 Apriori 算法获取典型焊接类型的算例,可以根据不同的材料属性和工艺要求获得不同焊接类型,同时给出了每条规则的支持度和置信度。该算法能较好地解决典型焊接类型获取问题。

基于关联规则的工艺提取,可以加强对工艺知识的重用,缩短新产品的开发周期,提高工艺规划的智能化水平,进而提高企业竞争力。因此,基于关联规则的工艺提取可以被广泛应用于工艺决策中。

参考文献

- [1] 韦卫星,莫赞,廖一奎.一种基于神经网络的知识获取方法研究与应用.计算机工程与应用,2004,40(5):95-98
- [2] 刘书暖,田锡天,张振明等.基于 Apriori 算法的典型工序序列获取方法.计算机集成制造系统,2006,12(8):1279-1283
- [3] 朱玉全,杨鹤标,孙蕾.数据挖掘技术.南京:东南大学出版社,2006,25-34
- [4] 刘书暖,张振明,田锡天等.基于聚类分析法的典型工艺路线发现方法.计算机集成制造系统,2006,12(7):996-1001
- [5] 刘闯,王俊彪.钣金 CAPP 中实例检索的灰色关联方法.机械科学与技术,2005,24(2):237-239
- [6] 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.机械工程学报,2004,40(5):843-847
- [7] Tsai C F, Lin Y C, Chen C P. A new fast algorithms for mining association rules in large databases. In: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Yasmine Hammamet, Tunisia, 2002. 251-256

The application of association rules to typical welding type acquisition

Chao Yongsheng, Liu Haijiang

(School of Mechanical Engineering, Tongji University, Shanghai 201804)

Abstract

To acquire the potential and valuable welding process knowledge from the process planning databases for body in white (BIW), the association rules were applied to capturing typical welding types. Firstly, an association rule model for mining typical welding types was established. Then the attributes affecting the welding types were classified, and among them quantitative attributes were discretized according to the welding process requirements. Secondly, the Apriori algorithm was used to extract frequent attribute item sets. Items belonging to the different attributes were joined to generate the candidate item sets. The strong rules whose consequents are welding type were generated according to the minimal confidence. Finally, welding data of a BIW were processed. To get desirable results the low support and the high confidence were applied. The generated rules had great reference value to welding type selection. This method can greatly improve the BIW process planning intelligence and efficiency.

Key words: association rules, Apriori algorithm, welding type, knowledge discovery, body in white, process planning