

改进的相似度算法在复合材料构件成型工艺检索中的应用

周玮珠, 梅中义

(北京航空航天大学机械工程及自动化学院, 北京 100191)

[摘要] 对比传统的局部相似度的计算方法,指出了其局限性,在考虑了生产实际情况中可能出现的几种情况后,提出了不同属性局部相似度计算的方法,使其能解决复合材料构件成型工艺系统中相似实例检索的问题,系统检索结果符合实际生产情况,表明了该方法在复合材料构件成型工艺相似实例检索中应用的可行性和准确性。

关键词: 复合材料构件; 成型工艺实例; 相似度计算; 实例检索; 不同属性局部相似度

Application of Improved Similarity Algorithm in Case Retrieval of Composites Molding Process

ZHOU Weizhu, MEI Zhongyi

(School of Mechanical Engineering & Automation, Beihang University, Beijing 100191, China)

[ABSTRACT] Compared to the traditional local similarity calculation method, it points out its limitations. Considering the several possible situation in actual production, it puts forward the partial similarity calculation method of different attribute to solve the problem of similar case retrieval in the aircraft composites component molding process system. The system retrieval results are in accord with the actual production. It indicates the feasibility and accuracy of this method in the application of case retrieval of the aircraft composite component molding process.

Keywords: Composites component; Molding process instance; Similarity calculation; Case retrieval; Partial similarity of different attributes

DOI:10.16080/j.issn1671-833x.2016.08.076

随着航空制造企业的发展,航空产品的成型工艺设计的经验也随之越来越丰富,而这些经验目前正以大量的实例形式被保存起来。如何有效地利用这些实例来提高复合材料构件工艺成型设计的效率无疑具有重要的意义。基于实例推理(CBR)是一种传统且有效的解决方法,而相似度计算是实例检索的关键技术。很多专家学者已经从不同角度对相似度计算进行了研究:蒋占四^[1]等给出了区间值属性相似度计算模型,并将各种属性类型的相似度计算方法加以统一;周宏明^[2]等建立了基于实体属性的相似度计算模型,提出了产品配置过程中相似度计算流程和方法;闫向彤^[3]等建立了向量空间模型,采用余弦函数值衡量两者间的相似程度。总体上看,这些方法均是基于数值型的用户需求参数和实例参数进行属性相似度计算。但很多时候某些属性无法采用数值型参数直接表示,那么其相应的属性相似度也无法求解。当然也有一些学者考虑了字符型的属性,如胡伟等^[4]提出了基于有向模糊图的产品系统结构相似度计算模型,但其只考虑了简单的情况。而对于复杂的复合材料构件成型工艺实例来说,这些方法都不能完

全适用,不符合企业的实际生产需求。

本文针对于某航空企业构建了复合材料构件成型工艺知识库系统,结合企业的实际需求,将实例推理应用于复合材料构件的成型工艺设计中,针对复合材料构件的特点,考虑了成型工艺实例中的材料信息,外形信息等不同属性的信息值,提出了一种基于复合材料构件成型工艺实例的不同属性局部相似度计算方法,结合改进后的相似度计算算法,得到与目标设计问题匹配度较高的解决方案,以便高效率、高质量地支持复合材料构件成型工艺的设计。

1 基于实例推理的知识组织形式

基于实例推理所使用的知识即为以往的成功实例,而这些实例都是以基于实例的知识模型存储在知识库中,它们均包含管理属性、问题状态描述属性以及解决方案3部分^[5]。复合材料构件成型工艺实例所包含的问题描述类详细信息主要为待设计构件的外形特征、设计要求以及生产要求等,而管理属性主要包括机型、零组件名称等信息,解决方案信息则包括所采用的成型工

艺、主要工艺参数以及工序信息等,其中工艺参数以及工序信息因成型工艺的不同而不同。那么,较为详细的复合材料构件成型工艺实例数据结构可用图1表示。

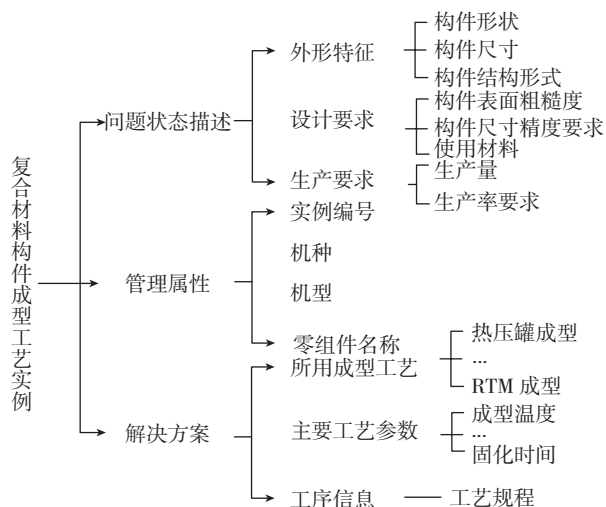


Fig.1 Data structure diagram of the composites molding process

系统进行推理计算时主要计算实例问题描述部分信息的相似性,当该部分信息相似时,即可判定两个实例相似,从而将成功实例的解决方案信息反馈给用户,用户可在该方案基础上进行修改,最终完成成型工艺设计。

2 常用相似度计算算法及其局限性

实例相似度是判别实例间相似性的尺度,是用来检索知识库中与待设计实例相似实例的机制。两个实例间相似度计算是通过计算其各个属性的局部相似度,在此基础上各自乘以一定的权重,最后将所有带权重的局部相似度累加,最终结果便是两个实例的相似度。整个相似度计算过程可以这样简单表示:实例 A 与实例 B 的相似度为 $\text{Sim}(A \cdot B) = \sum_{i=1}^n W_i \times \text{Sim}(P_{Ai} \cdot P_{Bi})$, 其中 P_{Ai} 表示 A 实例的第 i 个描述性属性, P_{Bi} 表示 B 实例的第 i 个描述性属性,而 w_i 表示实例的第 i 个描述性属性的权重,即表征了该属性在众多描述性属性中的相对重要程度(针对于评判相似度的情况),相似度 $\text{Sim}(x \cdot x)$ 的取值范围定为 $[0,1]$,值越大表示越相似。

可见实例相似度计算的关键即为局部相似度的计算,本文所分析的复合材料成型工艺实例的描述性属性有:构件形状、构件尺寸、构件结构形式、构件重要面表面粗糙度、构件尺寸精度要求、使用材料以及生产量,所以只需计算这些属性的局部相似度即可完成两个实例相似度的计算^[6]。一般的相似度计算方法为根据属性值类型的不同分为:

①当第 i 属性值为字符型

$$\text{Sim}(P_{Ai} \cdot P_{Bi}) = \begin{cases} 0, & V_{Ai} \neq V_{Bi} \\ 1, & V_{Ai} = V_{Bi} \end{cases}, \quad (1)$$

②当第 i 属性值为数值型

$$\text{Sim}(P_{Ai} \cdot P_{Bi}) = \frac{1}{1 + |V_{Ai} - V_{Bi}|}, \quad (2)$$

式中, V_{Ai} 、 V_{Bi} 分别为实例 A 、 B 第 i 属性的值。

然而,这种方法在对复合材料构件成型工艺实例的局部相似度进行计算时,显得非常不全面,不能直接应用上述相似度计算方法,因为仅仅通过判断字符值是否相同来断定某一属性是否相似是不切实际的,举个例子来看看上述方法存在的问题:构件使用材料,这一属性是通过材料牌号表示的,不同材料牌号的材料工艺性能不一定完全不同,事实上很多不同牌号的复合材料只是某些添加剂的比例不同而已,往往工艺性能也很相近,所以如果单纯的依据判断两个实例所使用的材料牌号字符串是否相同来计算材料属性的相似度会带来很大的计算误差。另外,在计算数值型属性的局部相似度时,上述公式(2)的计算结果有时也会不切实际,因为不同的属性的取值范围不同,而两个实例属性值差的绝对值大小并不能真实的反映相似程度,例如:构件尺寸精度取值范围为 $[1,18]$,而生产量取值范围 $[1,10000]$,量纲不同,如果单一的用公式(2)来计算,当两个实例尺寸精度分别为 1 和 18,生产量要求分别为 100 和 108 时,计算出这构件尺寸精度相似度和生产量要求相似度是相同的,然而在实际生产中,尺寸精度值为 1 和 18 完全可以判定为不相似,生产量为 100 和 118 时,可以判定为相似,所以公式(2)无法用同一相似度值来表征不相似与相似这两种完全相反的情况。由此可见,上述局部相似度的计算方法无法适用于不同量纲的、不同类型属性相似度的计算,针对这些问题,结合复合材料构件成型工艺实例的实际情况,设计了适用于复合材料构件成型工艺实例相似度计算的算法。

3 不同属性局部相似度计算方法的改进

复合材料构件成型工艺实例的描述性属性有:构件形状、构件尺寸、构件结构形式、构件重要面表面粗糙度、构件尺寸精度要求、使用材料以及生产量,其中构件形状、构件结构形式和使用材料属性为字符型属性较为特殊,构件尺寸、构件重要面表面粗糙度、构件尺寸精度要求和生产量为数值型属性,本文的相似度算法针对于不同的属性采取了不同的相似度计算方法。

(1)字符型属性相似度的计算方法。

·构件结构形式相似度计算

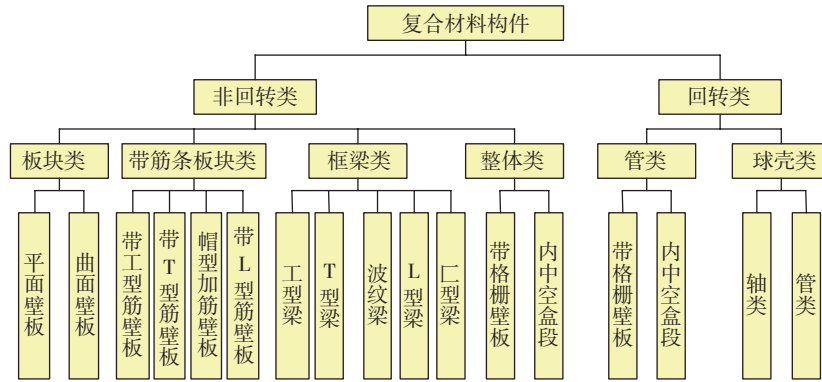


图2 复合材料构件外形形状分类图

Fig.2 Composites component shape classification

构件结构形式属性是表征构件内部构造的属性,当前企业中常用复合材料构件结构为壁板内部紧密压实的层压结构和内部用空隙的夹层结构,所以构件结构形式取值范围只有层压结构和夹层结构两种情况,该属性的相似度计算可以完全使用公式(1)计算。

· 构件形状相似度计算

构件形状是表征构件外形形状的属性,本文采用的知识库系统面向复合材料构件,所以只以复合材料构件作为研究范围。目前,复合材料广泛地应用于飞机结构件的制造,所以复合材料构件形状各异,将这些构件形状一一罗列是不现实的,美国麻省理工大学(MIT)根据外形特点将常用的复合材料构件分成一些基本形状^[7]。

通过与企业工程技术人员合作对企业自身使用情况的研究,在原有基础上做了一定的修改,如图2所示。借鉴对构件形状进行分类划分,构件形状属性相似度的计算可以设计为:首先判断两个构件形状属性值是否相同,相同则相似度为1,结束计算,不相同则判断两个构件形状属性值是否属于同一二级分类,如果属于同一二级分类,则相似度为0.7,不属于则继续判断是否属于同一一级分类,如果属于则相似度为0.4,不属于则相似度为0。

· 使用材料相似度计算

一个复合材料构件的设计信息中必然包含其所使用的材料信息,因为复合材料构件成型过程与材料性能有很大的关系,工艺人员进行成型工艺设计就必须获得该构件的材料信息,在实际生产中,构件使用材料是用材料牌号表征,而复合材料构件与金属构件不同,一个飞机构件可能有多种材料构成,通常设计人员为了使构件满足不同的力学、物理性能,会利用2~3种预浸料作为构件的铺层材料。那么在计算两个构件使用材料属性的相似度时,就不能简单的通过比较两个实例的使用材料牌号是否相同来计算。在工艺设计人员看来,两种材料相似只需它们的工艺性能相似即可,并不需要它

们的力学性能、物理性能等也相似,所以从材料的工艺性能入手计算两种材料的相似度^[8]。复合材料的工艺性能主要包括:成形温度、成形时间、凝胶时间、铺覆性、黏度和固化收缩率。两种材料的相似度本质也是通过计算其各个工艺性能属性的局部相似度得到的,本系统的两个实例使用材料集相似度计算流程为:

①分别获取待设计构件与成功实例的使用材料集 $M_A = \{m_{A1}, m_{A2}, \dots, m_{An}\}$ 与 $M_B = \{m_{B1}, m_{B2}, \dots, m_{Bn}\}$, 其中 M_{Ai} 、 M_{Bi} 分别代表实例 A、B 所使用材料的集合, m_{Ai} 、 m_{Bi} 即为单种材料;

②判断 M_A 是否属于 M_B , 如果属于则表示 A 的材料, B 实例中都有,所以应用于 B 实例材料的工艺参数也可以适用于 A 实例的材料,那么 A、B 实例使用材料相似度为 1, 结束计算, 如果 M_A 不属于 M_B , 则转入步骤③计算;

③判断 M_A 中是否有材料属于 M_B , 即是否存在 $m_{Ai} = m_{Bj}$, i, j 可以不相等, 如果存在则转入步骤④计算, 如果不存在则转入步骤⑤计算;

④记录 M_A 与 M_B 相同材料为集合 M_S , 由于 M_S 的材料在 M_B 中都存在, 所以 M_S 中每种材料 m_j 与 M_B 的相似度为 1, 这里记为 $sim_{m_{Aj}} = 1$, 而 $M_L = M_A - M_B$ 中的每种材料与 M_B 的相似度通过步骤⑤计算;

⑤材料集 M_X 、 M_Y 中不存在相同的材料时, 并不代表 M_X 中的材料与 M_Y 没有丝毫相似之处, 欲计算 M_X 中的某种材料与 M_Y 的相似度, 则需通过步骤⑥——计算出该种材料与 M_Y 中材料的相似度, 然后取最大值作为该种材料与 M_Y 中材料的相似度, 例如: M_X 中某一材料为 m_{Xi} , 经过利用步骤⑥——计算它与 M_Y 中各材料的相似度为 $sim(m_{Xi} \cdot m_{Yh}) = 0.3$ 、 $sim(m_{Xi} \cdot m_{Yk}) = 0.7$ 、 $sim(m_{Xi} \cdot m_{Yj}) = 0.5$, 则材料 m_{Xi} 与 M_Y 中材料的相似度为 $sim_{m_{Xi}} = 0.7$;

⑥现设 M_X 中某材料 m_{Xk} , M_Y 中的一材料 m_{Yn} , 两种牌号不同的材料的相似度是通过计算它们的工艺性能

的局部相似度得到的,即 $sim(m_{xk} \cdot m_{yh}) = \sum (w_{xx} \times sim_{xx}(m_{xk} \cdot m_{yh}))$,其中 xx 表示复合材料工艺性能属性,取值范围为[成形温度,成形时间,凝胶时间,铺覆性,黏度,固化收缩率], w_{xx} 是代表相应工艺性能属性相对于其他属性在材料成型时重要程度, $sim_{xx}(m_{xk} \cdot m_{yh})$ 则是表示 m_{xk} 、 m_{yh} 的局部相似度,由于材料工艺性能属性值均为数值型,所以该局部相似度的计算利用公式(2)中数值型属性相似计算方法进行计算的;

⑦经过步骤④、⑤的计算,可以得到 M_A 中每种材料与 M_B 中材料的相似度,下面只需计算 M_A 材料集与 M_B 材料集的相似度,该相似度计算公式如下:

$Sim(M_A \cdot M_B) = \sum_{i=1}^{n_{all}} (W_i \times sim_{mi})$,其中 n_{all} 为 M_A 中材料种数, sim_{mi} 分别由④、⑤步骤得到, W_i 表示材料 m_i 占整个构件构成的比例,计算方法见步骤⑧;

⑧一个实例中的某种材料与另一个实例材料的相似度只能表征它所构成部分与另一构件的相似度,所以必须乘以其在整个构件中的构成比,材料 m_i 在实例 A 中的构成比即为 W_i ,其计算公式为: $W_i = n/n_{all}$, n 表示实例 A 使用材料 m_i 的铺层数, n_{all} 表示实例 A 总铺层数;

通过上述①~⑧步骤完成了两个实例材料集相似度的计算,该计算方法从实际的生产情况出发,通过计算材料工艺性能属性的局部相似度,实现了不同材料集、不同材料牌号的相似度计算,计算结果完全符合实际成型工艺设计需求,解决了之前材料相似计算方法的结果误差大,不能进行材料集相似计算的问题。

(2) 数值型属性相似度的计算方法。

复合材料构件成型工艺实例的数值型描述性属性有:构件尺寸、构件重要面表面粗糙度、构件尺寸精度要求以及生产量,其中构件尺寸表示构件的最大尺寸,其取值范围为0~15m,构件重要面表面粗糙度取值范围则是0.12~50,构件尺寸精度取值范围则为1~18,而生产量取值范围则为1~1000,这些属性的量纲不同,而又需使它们的相似度计算结果在0~1的范围内,所以必须对这些不同量纲的属性值进行归一化处理。一般的归一化处理函数为: $g(x) = |x_1 - x_2| / (x_{max} - x_{min})$,其中 x_1 、 x_2 为某属性待计算相似度的两个值, x_{max} 、 x_{min} 分别为该属性的最大、最小值,即取值范围^[9]。

但是,在实际生产中,这样的归一化处理方式仍然有一定的不足,因为各属性的取值范围值,是专家依靠以往的制造经验预估的大致取值范围,并不能保证在企业以后的生产中,某些属性值会超出现有取值范围,当有属性值超出取值范围时,该归一化处理方式将无法计算,不利于企业知识库系统的长久使用,要消除这一弊端就必须

将属性值超出取值范围的情况考虑在内,在原有归一化处理函数的基础上增加了一些处理意外情况的函数,就构成了一个归一化处理的分段函数,具体如下:

$$g(x_1, x_2) = \begin{cases} \frac{|x_1 - x_2|}{x_{max} - x_{min}}, & x_{min} \leq x_1 \leq x_{max} \text{ 且 } x_{min} \leq x_2 \leq x_{max} \\ 0, & x_1 \geq x_{max} \text{ 且 } x_2 \geq x_{max} \\ \frac{|x_1 - x_2|}{x_{max} + x_{min}}, & \text{其他} \end{cases} \quad (3)$$

这样的处理方式是在充分考虑了工程实际中专家认知基础上完成的,当两个被计算值都在取值范围之内时,则按一般的归一化处理方式处理,而当两个属性值都超出取值范围时,可直接视它们没有差别,所以直接将它们差值视为0,例如:若两个实例生产量值分别为1001与1100时,在工程技术人员看来这都是属于大批量生产,在工艺条件的设计和模具设计时都会按照大批量生产的角度进行,所以可直接视它们的差值为0。而当两个属性值一个在取值范围中,另一个超出取值范围时,则利用 $\frac{|x_1 - x_2|}{x_{max} + x_{min}}$ 对它们的差值进行归一化处理,这样处理可以在考虑生产实际的基础上避免计算结果超出归一化的范围,例如:当生产量 $x_1=1$, $x_2=1010$ 时,如果还按照 $\frac{|x_1 - x_2|}{x_{max} - x_{min}}$ 进行归一化处理,则计算结果大于1,违反了归一化处理的初衷,而利用 $\frac{|x_1 - x_2|}{x_{max} + x_{min}}$ 处理后,结果完全在0~1的范围内,并且真实反映了生产实际情况,生产量为1属于单件生产的范畴,而生产量1010属于大批量生产的范畴,所以它们差值应该是接近1的,计算结果符合该生产实际。经过上述归一化处理,不同量纲的数值型属性值的差值均处在0~1的范围内,则两个属性值相似程度反映为差值越小越相似,然而总体的相似程度表征是相似程度越大,相似程度越大,所以必须对上述归一化处理的结果进行再处理使其符合相似程度表征的规律。构造一个分段函数对归一化处理结果进行了再处理,详细函数表达式如下:

$$f(g(x_1, x_2)) = \begin{cases} e^{-g(x_1, x_2)}, & 0 \leq g(x_1, x_2) \leq -\ln a \\ \frac{a(1 - g(x_1, x_2))}{\ln a + 1}, & -\ln a < g(x_1, x_2) \leq 1 \end{cases} \quad (4)$$

其中, a 为常数,取值范围为(0,1),如图3所示。

a 为调节因子,即相似计算的一个阈值,阈值以下部分的相似程度值较小,对最终相似度计算影响较小,所以在阈值以下部分采用一次函数快速逼近于0。综上所述,采用式(3)、(4)解决了复合材料构件成型工艺实例数值型属性量纲不同、取值范围不固定的问题,同时该相似度计算方法充分考虑了实际生产情况,符合领域专家认知。

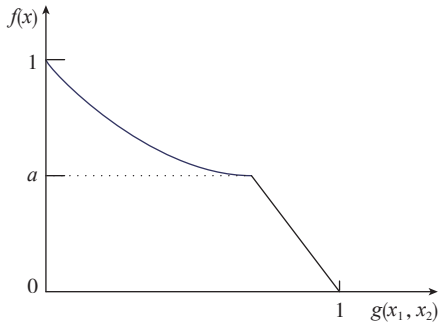


图3 数值型属性相似度计算函数图像

Fig.3 Picture of the numeric attributes similarity calculation function

4 应用实例

现利用所开发的系统对一实例进行成型工艺相似实例的检索,系统操作界面如图4所示。

实例参数如下:T型梁,生产10件,最大平面尺寸为1500mm,表面粗糙度要求为12.5,精度要求是6,使用材料型号为759/3233,铺10层。

根据以上阐述的方法进行局部相似度计算,复合材料构件成型工艺实例的描述性属性中的构件形状、构件结构形式和使用材料属性为字符型属性,构件尺寸、构件重要面表面粗糙度、构件尺寸精度要求和生产量为数值型属性。

构件结构形式分为层压结构和夹层结构,有相同和不同两种情况,对应值分别为<1,0>。构件形状表示为完全相同,属于同一二级分类,属于同一一级分类,完全不同,对应值分别为<1,0,7,0,4,0>。使用材料同样为字符型因素,虽然该因素也可通过树脂类别进行分类,但是同类树脂工艺性能仍然存在很大不同,所以不能按照类似于制件形状因素的计算方法计算相似度。材料的工艺性能的影响因素很多,例如材料组成,固化温度、成形压力、树脂黏度和固化收缩率等,这些因素有的为字符型,有的为数值型,所以可以利用改进的局部相似度计算方法,根据上述材料相似度计算流程,以材料的属

性为因素集,计算其对应每种成型工艺实例的材料相似度。数值型的属性相似度计算根据公式(3)和(4)进行计算。在检索过程中,不同的用户有不同的需求,故系统可根据用户的需求,给每个属性配以一定的权重,最后得到计算结果。表1中显示总相似度大于0.5的实例。

表1 复合材料构件工艺成型实例详细相似度计算结果

实例编号	形状相似度	产量相似度	尺寸相似度	表面相似度	精度相似度	材料相似度	总相似度
1	1.0000	1.0000	0.8500	0.5016	1.0000	1.0000	0.9751
2	0.7000	1.0000	1.0000	0.5016	0.9000	1.0000	0.8551
3	1.0000	0.7800	1.0000	0.5016	0.9000	0.3552	0.7494
4	0.4000	0.7800	1.0000	0.5016	0.9000	1.0000	0.7351
5	0.7000	0.7800	0.7800	0.5016	1.0000	0.3552	0.6294

根据最优原则,应该选择总相似度较高的实例为参考实例。

5 结论

本文论述的改进后的相似度检索算法应用于复合材料构件成型工艺知识库系统中,其检索结果符合生产实际情况,由此可得以下结论:

(1)改进后的局部相似度检索算法完全能通过软件系统实现;

(2)该相似度计算算法充分考虑了复合材料构件成型工艺的属性信息,如飞机构件所采用的材料等非常规类字符属性。由于复合材料的特殊性,不同名称的材料不一定完全不相似,故提出了复合材料构件的材料相似度计算流程及方法,符合生产实际需求;

(3)复合材料构件生产中很多数值型属性,它们属于不同量纲,故需将其归一化,但在实际生产中,某些属性值会超出其限定的取值范围,传统的归一化处理无法计算这些意外的情况。改进后的相似度计算算法充分地考虑了实际生产情况,综合处理了各种意外情况,提高了算法的适应性,将字符型属性和数值型属性的相似度计算进行了全面的考虑,从而使计算结果可靠可信。

总体而言,通过将改进后的相似实例检索算法应用到计算机辅助系统,大大提高了复合材料构件成型工艺检索的效率和准确率。

参考文献

[1] 蒋占四,陈立平,罗年猛.最近邻实例检索相似度分析[J].计算机集成制造系统,2007,13(6):165-168.

JIANG Zhansi, CHEN Liping, LUO Nianmeng. Similarity analysis

(下转第84页)



图4 复合材料构件成型工艺实例检索界面

Fig.4 Retrieval interface of composites component molding process case