

DOI:10.3969/j.issn.1001-4551.2017.02.001

机械结构对称性实例设计知识挖掘方法研究^{*}

杨俊成,马志勇^{*},张太霞

(1. 宁波大学 机械工程与力学学院,浙江 宁波 315211;2. 浙江省零件轧制成形技术研究
重点实验室,浙江 宁波 315211)

摘要:针对机械结构对称性实例中存在着大量的设计知识,利用多维模糊数据挖掘的方法从结构对称性实例中挖掘了设计知识,用以指导结构对称性在机械产品结构设计中的科学应用。提出了根据设计需求结合隶属函数特征对实例数据库中待挖掘的数据模糊化处理,借助关联规则挖掘方法 Apriori 算法对预处理后的数据库进行了挖掘。挖掘出了结构对称性和产品功能区间、性能区间及约束区间之间的模糊关联规则。结果表明:通过对结构对称性实例进行数据挖掘,建立结构对称性在机械系统中实用的设计知识,是以后进一步研究机械结构对称性应用规律和应用方法的坚实基础。

关键词:机械结构对称性;数据挖掘;隶属函数;模糊关联规则;设计知识

中图分类号:TH122;TH166

文献标志码:A

文章编号:1001-4551(2017)02-0105-06

Data mining method for design knowledge in mechanical structure symmetry instances

YANG Jun-cheng, MA Zhi-yong, ZHANG Tai-xia

(1. Faculty of Mechanical Engineering & Mechanics, Ningbo University, Ningbo 315211, China;
2. Zhejiang Provincial Key Lab of Part Rolling Technology, Ningbo 315211, China)

Abstract: Aiming at a lot of design knowledge in mechanical structure symmetry instances, by using the multidimensional fuzzy association rules mining method, design knowledge was mined from structure symmetry instances to guide the scientific application of structure symmetry in mechanical product design. The data to be mined in mechanical structure symmetry instance database was processed by fuzzy method based on the design requirements and the characteristics of the membership function, and the association rules mining algorithm Apriori was used to mine the fuzzy association rules. The mining results include the relationships among the structural symmetry, the mechanical product function, the mechanical product performance and the design restriction. The results indicate that the practicable structural symmetry design knowledge in mechanical systems is established by mining the structure symmetry instances, and the reference of further research on the application rules and application methods of mechanical structure symmetry.

Key words: mechanical structure symmetry; data mining; membership function; fuzzy association rules; design knowledge

0 引言

结构对称性是机械产品结构中的一种比较常见的

现象,有效采用数据挖掘方法从结构对称性实例数据库中挖掘出机械结构对称性与对称性效用之间的关联规则,建立结构对称性设计知识体系,有助于指导结构

收稿日期:2016-10-10

基金项目:国家自然科学基金资助项目(51305213)

作者简介:杨俊成(1989-),男,江苏淮安人,硕士研究生,主要从事结构对称性设计知识挖掘方面的研究. E-mail: nbuyjc@163.com

通信联系人:马志勇,男,讲师,硕士生导师. E-mail: mazhiyong@nbu.edu.cn

对称性在机械产品设计中的科学应用,提高机械产品的技术性、经济性和社会性^[1]。在现有的研究中,对机械结构对称性设计知识挖掘及提炼已有初步的研究,如浙江大学冯培恩课题组^[2-3]已建立了以关联规则为基础的结构对称设计知识模板、设计知识挖掘方法等,但是现有的设计知识挖掘方法比较简单,而且主要集中在机构对称,并不能反映对称破缺、弱对称相关的设计知识,由此得到的挖掘结果不够深入、完整。本研究旨在理论分析和实例分析的基础上,提取若干条设计规则,用以完整地挖掘结构对称性设计知识。

自 Agrawal 等人^[4]于 1993 年提出经典的挖掘数据库中项集间关联规则挖掘方法以来,国内外针对关联规则的数据挖掘方法已进行了大量的研究,总结了大量的设计知识应用规律。Delgado 等人^[5]提出了挖掘模糊关联规则的一般模型,并给出了模糊关联规则模型在多媒体数据挖掘、Web 挖掘等具体领域的应用。Latiri 等人^[6]通过应用模糊关联规则查询扩展的方式来提高信息检索系统的查全率以及查准率;Lopez 等人^[7]将模糊关联规则应用于酵母的生物数据分析研究;Delgado 等人^[8]将模糊关联规则用于对橄榄耕种测量资料的评价;王婷等人^[9]在关联规则挖掘研究的基础上,利用数据挖掘和国内互联网结构特征,提出一种 IP 定位方法,拓展了数据挖掘在互联网领域的应用等等。

现有的研究已将关联规则挖掘方法应用在多个领域,但目前在机械结构对称性领域尚缺少比较系统的研究。本研究主要采用频繁项集挖掘方法,对结构对称性实例数据库中支持度和置信度较高的关联规则进行总结和提炼。在多维模糊关联规则挖掘过程中,通过引用隶属函数可以有效解决区域划分边界过硬的缺点,挖掘出丰富的实用的设计知识。笔者根据不同层次分别抽象出不同的应用规律,总结出不同深度、不同适用范围的应用知识,并根据实例库中的关联规则的提取,总结出若干条设计规则,将关联规则应用到实例的结构设计、机构选型等多领域中,以指导设计人员科学地创新设计。

1 多维模糊关联规则挖掘过程

马志勇、刘帆等^[10]已将收集的实例标准化,并建立实例数据库,本研究主要对实例数据库中结构对称性与对称类型实现的设计需求之间的关联规则(结构对称性 $\Rightarrow$ 设计需求)进行数据挖掘。在实例数据库中,每一对称性实例会有对应的功能、效用和约束。结

构对称性设计实例与设计需求常见的 3 种映射过程如图 1 所示。

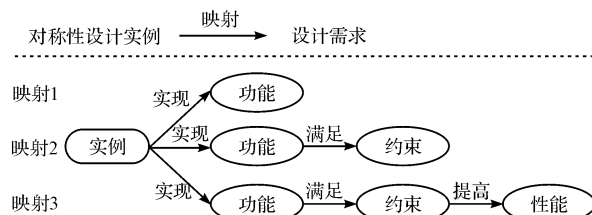


图 1 对称性结构与设计需求关系

BOM 表记录了结构对称性实例数据库中结构对称类型与设计需求的关联知识,并用支持计数表示各个属性的隶属度程度。在一般的数据挖掘过程中,支持度和置信度的阈值被用于表示关联规则的相关性。支持度用来确定规则可以用于给定数据集的频繁程度,反映发现的规则的有效性;置信度用来确定 B 在包含 A 的事务中出现的频繁程度,反映发现的规则的确定性^[11],且支持度和置信度的数学模型为:

$$\text{support}(A \rightarrow B) = P(A \cup B) = \frac{|\{T \mid T \in D \wedge (A \cup B) \subseteq T\}|}{|D|} \quad (1)$$

$$\text{confidence}(A \rightarrow B) = P(B \mid A) = \frac{|\{T \mid T \in D \wedge (A \cup B) \subseteq T\}|}{|\{T \mid T \in D \wedge A \subseteq T\}|} = \frac{\text{support}(A \cup B)}{\text{support}(A)} \quad (2)$$

式中: T, D —事务的集合。

机械结构对称性与设计需求之间的设计规则需要考虑实例数据库中的数据维数。单维关联规则涉及处理不同实例同种属性的关系,例如结构对称实现功能属性的关系;多维关联规则主要是从实例数据库中挖掘实例不同属性的关系^[12],例如结构对称在实现功能区间属性、性能区间属性和约束性能属性之间的关系。具体挖掘方法需要考虑数据项集间的属性性质,项集之间的约束,并且每一维都需要人为地设置支持度和置信度阈值。

在提取关联规则的过程中,支持度和置信度的阈值的确定对挖掘效果的影响较大,而且并非唯一确定,需要在挖掘过程中反复修正。为了获得较为合理、比较有意义的关联规则,结合研究的目的要求,本研究根据专家系统意见,进行科学取值。结构对称性设计知识挖掘主要以对称性、设计需求等为出发点,对结构对称性实例库进行多角度、多层次的知识挖掘,可为实例数据库中的具体实例的应用提供可靠依据。结构对称性设计知识模糊关联规则挖掘包括机械结构对称性设计实

例库中实例的数据准备、数据挖掘以及结果解释和评估的过程^[13],主要挖掘过程如图2所示。

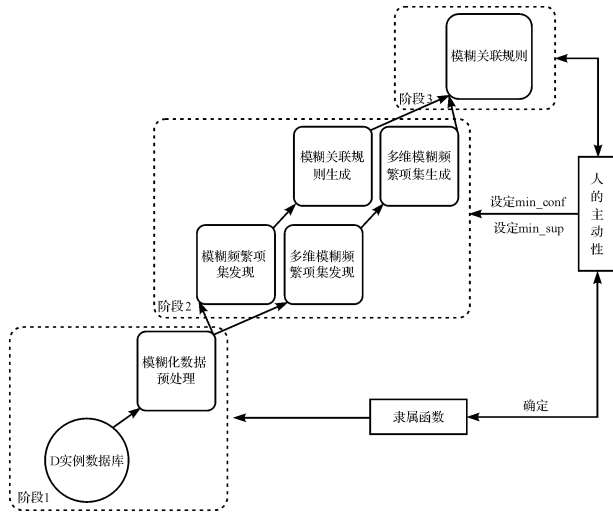


图2 多维模糊数据挖掘模型

1.1 数量型属性离散化

在实例数据库中,实例出现的概率表达了精确的概念,但功能区间、性能区间以及约束区间之间的关联是一个比较模糊的概念,因此需要引入模糊关联规则概念。用模糊概念表示数据库中实例属性之间的关系,通过采用模糊集的方法对实例的属性值进行转换,可以有效避开由于区间划分引起边界过硬的缺点。确定结构对称性实例数据库中数据的隶属函数,首先需要对原数据进行数据处理,将分析的对象作为样本,如: $a_{i1}, a_{i2}, \dots, a_{ik}$,称 $A_{ik} = (a_{i1}, a_{i2}, \dots, a_{ik})$ 为样本集。这里把实例库中的属性数量化, a_{ik} 表示实例数据库中第 i 行样本 a_i 对第 k 项指标出现的概率,对于 k 维实例库的表示,即 $A_{ik} = (a_{i1}, a_{i2}, \dots, a_{ik}) (k = 1, 2, \dots, m)$,标准化处理样本数据的数学模型如下式所示:

$$a'_{ik} = \left| \frac{a_{ik} - \mu_k}{\sigma_k} \right| (k = 1, 2, \dots, m) \quad (3)$$

式中: $\mu_k = \sum_{i=1}^n a_{ik} (k = 1, 2, \dots, m) / n$; $\sigma_k =$

$$\sqrt{\sum_{i=1}^n (a_{ik} - \mu_k)^2 / n} (k = 1, 2, \dots, m)。$$

根据实例数据库中数据的特性和实际计算简化得到的结果,并不能将数据控制在 $[0, 1]$ 区间内,增加了计算的复杂性,这里引用极差运算对得到的数据简化处理。极差运算的数学模型如下式所示:

$$a''_{ik} = \left| \frac{a'_{ik} - \min_{1 \leq k \leq n} \{a'_{ik}\}}{\max_{1 \leq k \leq n} \{a'_{ik}\} - \min_{1 \leq k \leq n} \{a'_{ik}\}} \right| (k = 1, 2, \dots, m)$$

(4)

聚类区间和隶属函数的确定在模糊属性的分析中具有重要的作用,而隶属函数的确定往往根据人们对模糊集的认识程度、实际经验等因数而不同,具有一定的主观性。综合以上因素,本研究选取三角形隶属函数,每个数量型属性选用3个聚类区间,隶属函数确定的模糊集与对称性实例数据库中数据属性分别映射为“区域 E , 函数 $f_1(a''_{ik})$; 区域 I , 函数 $f_2(a''_{ik})$; 区域 T , 函数 $f_3(a''_{ik})$ ”3个模糊域,其中:区域 E —实例数据库中实例出现的概率小,区域 I —实例数据库中实例出现的概率一般,区域 T —实例数据库中实例出现的概率高。对已收集的实例数据源进行扫描,用隶属度代替每个实例对应属性出现的概率值。隶属函数 F 如图3所示。

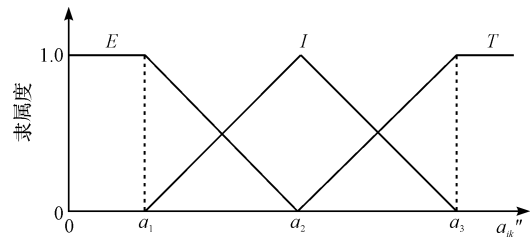


图3 模糊集合隶属函数 F

图3中, a_1, a_2 段分别由 $f_1(a''_{ik}), f_2(a''_{ik})$ 函数确定, a_2, a_3 段分别由 $f_2(a''_{ik}), f_3(a''_{ik})$ 函数确定。在模糊化的过程中,每个属性有3个聚类区间,需要将每个属性归一化,同时还要满足非负性条件,即 $\sum_{i=1}^3 f(a''_{ik}) = 1, f(a''_{ik}) \geq 0$ 。隶属函数 F 的数学模型为:

$$F(a''_{ik}) = f_1(a''_{ik}) + f_2(a''_{ik}) + f_3(a''_{ik}) \quad (5)$$

其中:

$$f_1(a''_{ik}) = \begin{cases} 1 & 0 \leq a''_{ik} < a_1 \\ \frac{a_2 - a''_{ik}}{a_2 - a_1} & a_1 \leq a''_{ik} < a_2 \\ 0 & a_2 \leq a''_{ik} \end{cases}$$

$$f_2(a''_{ik}) = \begin{cases} 0 & a''_{ik} < a_1 \\ \frac{a''_{ik} - a_1}{a_2 - a_1} & a_1 \leq a''_{ik} < a_2 \\ \frac{a_3 - a''_{ik}}{a_3 - a_2} & a_2 \leq a''_{ik} < a_3 \\ 0 & a_3 \leq a''_{ik} \end{cases}$$

$$f_3(a''_{ik}) = \begin{cases} 0 & a''_{ik} < a_2 \\ \frac{a''_{ik} - a_2}{a_3 - a_2} & a_2 \leq a''_{ik} < a_3 \\ 1 & a_3 \leq a''_{ik} \end{cases} \quad (6)$$

1.2 主要挖掘方法

(1) 本研究经过隶属函数模糊处理实例数据库中的数据,完成实例数据库中的数据初始化。将原有的实例数据库 a_{ik} 离散化,转换成新的实例数据库 a'_{ik} , a'_{ik} 的模糊属性与初始化的过程中维数相关,需要考虑实例数据库的维数问题:

(a) 当采用单维关联规则方法时,结构对称实现不同功能属性之间的关系,通过隶属函数将数据区间模糊化。采用 Apriori 算法,将 Apriori 算法和包含模糊属性的事务相结合,挖掘出感兴趣的规则形式。算法第一次迭代时,得到所有的 1- 模糊属性集的模糊概率,即候选 1- 模糊频繁属性集 C_1 。

(b) 当采用多维关联规则方法时,联系每一项的属性,将每一项的祖先加入到关联规则中。用 Apriori 算法扫描扩展的数据集^[11],产生候选项集。如果扩展项集中的项集大于预定义的最小支持度,则该项集是频繁的,否则,不是频繁的。在多维关联规则实例数据库中,机械产品的设计需求可以分为功能、性能和约束 3 种类型;对称结构主要分为对称、对称破缺和弱对称,对称结构中又分别包含平移、旋转和镜射等基本运动。通过这样的层次关系,进行第一次算法迭代,得到所有的 1- 模糊属性集的模糊概率,即候选 1- 模糊频繁属性集 C'_1 。

(2) 本研究设置最小支持度 $\min_sup$ 和最小置信度 $\min_conf$,满足支持度和置信度阈值的规则称为强关联规则。扫描实例数据库,删除不符合最小支持度

$\min_sup$ 的项,得到 1- 模糊频繁属性集的集合 L_1 。

(3) 为了得到 2- 模糊频繁属性集的集合 L_2 ,本研究通过 $L_1 \bowtie L_1$ 计算产生候选 2- 模糊频繁属性集的集合 C_2 ,扫描候选频繁项集集合 C_2 ,删除不符合最小支持度 $\min_sup$ 的项,得到 2- 模糊频繁属性集的集合 L_2 。

(4) 同理,本研究得到候选 2- 模糊频繁属性集的集合 C_3 ,删除不符合最小支持度 $\min_sup$ 的项,得到 2- 模糊频繁属性集的集合 L_3 。

(5) 本研究按照这样的方法执行,直至没有 k - 模糊频繁属性集产生为止。此时, $C_k = \phi$, Apriori 算法停止。

(6) 根据设计需求,本研究从模糊频繁属性集中提取有意义的关联规则。

在数据预处理的过程中,由于功能区间、性能区间、约束区间的重要性不同,即功能区间 > 约束区间 > 性能区间,需要分别对 3 个区间加权,令权数组成的集合 $\beta = (x_1, x_2, x_3)$ 。

2 实例中的推理规则

本研究建立的模糊关联规则挖掘方法用以总结结构对称性的应用知识及应用规律,挖掘出结构对称类型与对称类型能够实现的设计需求之间的关联规则(结构对称设计需求)。样本的数量决定规则提取的可靠性,为确保实例更有说服力,本研究采用了前期收集的 500 个实例进行知识挖掘^[14-18],结构对称性类型与设计需求关联知识如表 1 所示。

表 1 结构对称性类型与设计需求关联知识

对称性结构		设计需求				
		平移传送物料	双向停止转动	……	有害作用自抵消	扩大功能作用范围
对称	平移对称	6.14%	1.32%	……	9.68%	16.23%
	旋转对称	1.68%	6.97%	……	0.72%	6.01%
	镜射对称	1.97%	3.55%	……	8.28%	4.54%
……						
对称破缺	……					
弱对称	……					
……	……					

表 1 中,结构对称性类型与对称性类型能够实现的设计需求之间的关联知识,得到各种不同的对称性下实现的不同的设计需求的关联知识,表达了比较精确的概念,但实际上功能区间、性能区间、约束区间是一个比较模糊的概念,采用上述讨论的数据预处理方法,对称结构下 3 大基本运动形式的设计需求进行知识挖掘涉及的设计需求较多,本研究从表 1 中提取部分设计需求,对提取的设计需求属性进行数据预处理,

得到的模糊实例数据库如图 4 所示。

在图 3 所示的模糊集合隶属函数中,考虑实例数据库中数据本身模糊性,以及模糊函数的特征,结合专家意见,经过反复计算,当权值取 $x_1 = 0.1$, $x_1 = 0.5$, $x_1 = 0.9$ 时,得到的模糊关系较为合理。

候选项集和频繁项集计算的过程中,本研究把每个属性值离散化,设置最小支持计数为 0.9,将每一次扫描产生的候选频繁项集分别与最小支持计数 0.9 进

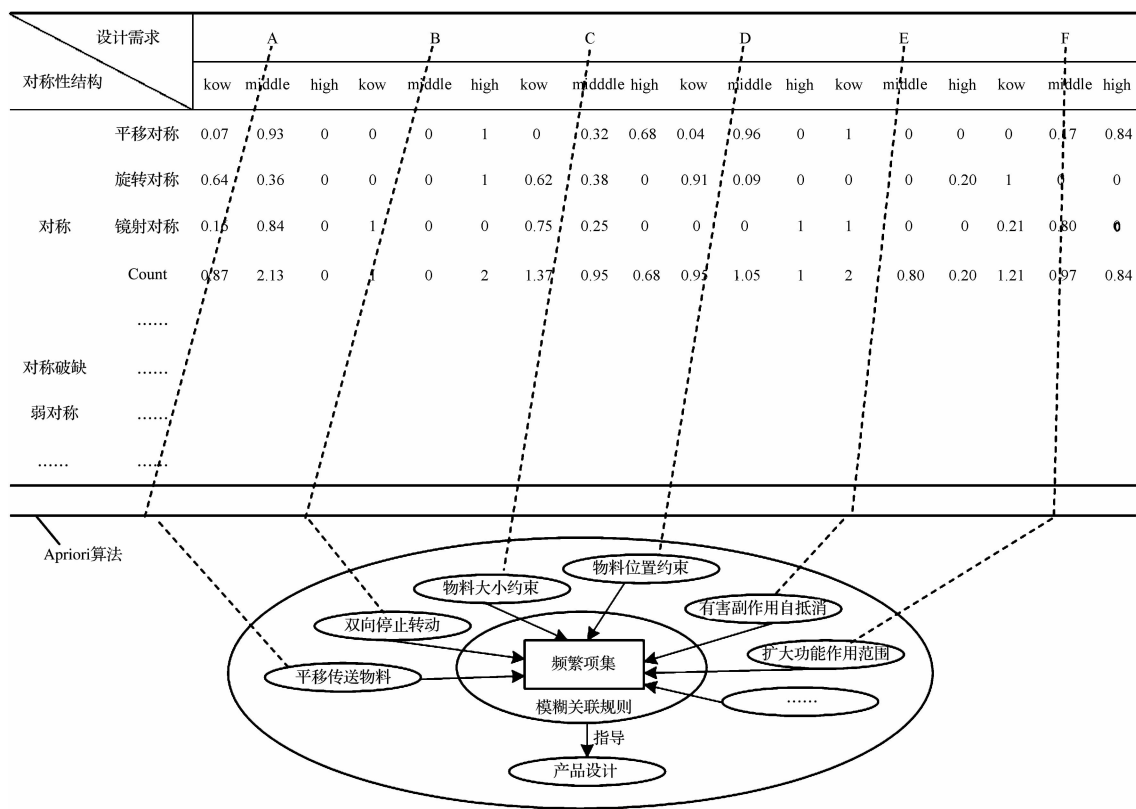


图4 模糊数据库 D

行比较,产生模糊频繁属性集,直至 $\text{count}(C_x) \leq 0.9$ 得到模糊频繁项集 $L_x = \phi$ 。模糊频繁集挖掘结束。

在得到的所有的频繁项集中,设 $\min_sup = 0.044$, $\min_conf = 0.7$,挖掘出若干条频繁模糊项集,由于涉及的关联规则较多,本研究仅列举出部分满足要求的关联规则,输出关联规则如下:

- 1) $\{A. \text{middle}, B. \text{high}\} \Rightarrow \{D. \text{middle}\}$,
support = 0.051 4, confidence = 0.717 2;
- 2) $\{B. \text{high}, D. \text{middle}\} \Rightarrow \{A. \text{middle}\}$,
support = 0.051 4, confidence = 0.804 5;
- 3) $\{B. \text{high}, F. \text{low}\} \Rightarrow \{D. \text{low}\}$,
support = 0.050 6, confidence = 0.91;
- 4) $\{A. \text{middle}, B. \text{high}\} \Rightarrow \{E. \text{low}\}$,
support = 0.051 7, confidence = 0.720 9。
- 5)

可以发现:当实例库中的实例出现中等概率的平移传送物料功能和高概率的双向停止转动功能→实例

出现中等概率物料位置约束规则中,三者同时出现的概率是 5.14%;并且在大量实例的实例数据库中,当有中等概率的平移传送物料功能和高概率的双向停止转动功能时,有 71.72% 的情况实例中具有中等概率物料位置约束,同理分析出其他的模糊关联规则。

3 结束语

本研究主要采用多维模糊关联数据挖掘的方法发现机械结构对称性实例数据库中功能区间、性能区间及约束区间之间的关联规则设计知识,该知识挖掘的结果有助于指导结构对称性在机械产品设计过程中的科学应用,该挖掘的方法可为其他领域的设计知识挖掘提供借鉴。本研究根据实例数据库不同的维度抽象出不同的应该规律,将综合模糊挖掘方法应用到实例数据库中,总结不同类型不同深度的应用知识。如何全面系统地挖掘结构对称性设计知识,拓展本研究的挖掘方法在其他领域的应用是笔者下一步的主要工作。

(下转第 114 页)

本文引用格式:

杨俊成,马志勇,张太霞. 机械结构对称性实例设计知识挖掘方法研究[J]. 机电工程,2017,34(2):105-109,114.

YANG Jun-cheng, MA Zhi-yong, ZHANG Tai-xia. Data mining method for design knowledge in mechanical structure symmetry instances[J]. Journal of Mechanical & Electrical Engineering, 2017,34(2):105-109,114.

《机电工程》杂志;http://www.meem.com.cn