

文章编号:1005-0523(2019)02-0105-06

# 基于密度算法和 DSM 的模块划分方法

程贤福, 万冲, 邱浩洋, 万丽云, 周健

(华东交通大学载运工具与装备教育部重点实验室, 江西 南昌 330013)

**摘要:**针对产品模块化设计中模块划分的问题,考虑到实际过程中设计结构矩阵的多样性,提出一种基于密度算法和设计结构矩阵的模块划分方法,以设计结构矩阵描述零部件间的关联关系,通过 $n$ 维空间的欧式距离判断各零部件间的联系强弱,利用密度可达原则进行零部件聚类。根据算法参数选用的不同获得不同划分结果,引用模块度准则判别最优的聚类结果作为最终的模块划分方案。最后以汽车起重机上车部分为例说明模块划分过程。

**关键词:**模块化设计;模块划分;密度算法;设计结构矩阵;汽车起重机

**中图分类号:**TH122

**文献标志码:**A

产品的模块化设计是为应对现代设计的高效性、多样性和产品再制造能力而提出的一种设计方法<sup>[1]</sup>。模块划分是该设计方法的核心部分,合理的模块划分直接影响产品的性能、成本和效率<sup>[2]</sup>。产品的模块划分方法一般采用基于底部零件、自下而上的聚类分析<sup>[3]</sup>。现有的聚类方法已经逐渐从手工聚类向智能聚类转变,常见的聚类算法大体上可以归为基于划分的方法、基于层次的方法、基于密度的方法、基于网格的方法和基于模型的方法<sup>[4]</sup>。而针对产品设计提出来的聚类方法又有所不同。如,Algeddawy等<sup>[5]</sup>在设计结构矩阵为布尔矩阵的前提下采用了基于层次的聚类方法进行产品模块智能划分,但产品设计中并非都是布尔矩阵,故导致其应用受限。刘建刚等<sup>[6]</sup>提出了一种遗传算法和产品设计结构矩阵(design structure matrix, DSM)相结合的智能模块化运算,通过选择、交叉、变异等过程获取最优聚类结果,但由于算法所需设置的运行参数过多,要获得较好划分结果需要进行较多次的调试。马淑梅等<sup>[7]</sup>对传统模糊C均值算法进行改进,并针对机床开发了模块划分系统平台,但针对不同产品无法表现较好的适应性。Alex Rodriguez等<sup>[8]</sup>提出了一种基于密度峰值的聚类算法,该算法能够发现任意形状的簇并且具有良好的抗噪性,同时也可避免由于密度分布不均影响聚类效果。

通过在前人的研究基础上,考虑到设计结构矩阵有布尔和非布尔、对称和非对称等多种类型,基于设计结构矩阵表达零部件间的关联关系,利用密度可达原则将零部件进行聚类,以模块内部的高内聚度和模块之间的低耦合度为划分准则选择最优的聚类结果为最终划分方案,为后续进行产品模块之间的关联影响分析及模块实现的优先级提供了依据。该方法将设计结构矩阵与密度算法相结合,并利用划分准则进行判断,在针对零部件数量较多的产品进行设计时可以高效准确地获得所需的模块划分方案,提高模块划分的合理性。

## 1 模块聚类与划分

模块划分是基于产品各个零部件之间的关联关系强弱而进行区分归类的一种方法。针对零部件间关联关系的表达,现在广泛采用设计结构矩阵的方式,该方式不但可以直观观察各零部件的联系,同时又可以实

收稿日期:2018-09-27

基金项目:国家自然科学基金资助项目(51765019,71462007)

作者简介:程贤福(1975—),男,教授,博士,研究方向为设计理论与方法、模块化技术、大批量定制等。

现强弱关系的量化表达。它以矩阵的形式表达了各个零部件间的联接、功能和物理关联性,对产品开发过程进行并行规划<sup>[9]</sup>。故本文采用 DSM 作为零部件间关联关系的表达方式,利用密度算法与 DSM 结合进行智能聚类,使得关联关系较强的零部件组成一个集合,不同集合间的关联关系尽可能的低。

### 1.1 密度算法介绍

针对 DSM 进行聚类,目前采用的智能聚类方法中密度算法具有明显的优势。密度算法是根据区域内点的密度进行聚类的一种算法,相比一般基于距离的算法可以有效处理噪点和聚类形状的问题。要理解密度聚类算法,首先必须理解以下几点定义:

- 1) Eps 领域:以选定对象为中心,半径 Eps 内的区域;
- 2) 核心对象:该对象的 Eps 领域内存在大于等于 MinPts 数量的样本点;
- 3) 直接密度可达:如果某对象  $p$  在核心对象  $q$  的 Eps 领域内,则称对象  $p$  是对象  $q$  的直接密度可达;
- 4) 密度可达:如果存在一个对象链  $p_1, p_1, \dots, p_n, p_1=q, p_2=p$ , 对于  $p_i (1 \leq i \leq n), p_{i+1}$  是  $p_i$  的直接密度可达,则对象  $p$  是对象  $q$  的密度可达。

密度算法的目的便是通过确定 Eps 和 MinPts 两个参数,在对象集中找出核心对象,并针对核心对象通过直接密度可达的方式,迭代找出所有的密度可达对象聚成一个簇,直到遍历所有对象后终止,并把无法聚类的对象定义为噪音点。由于密度算法是通过密度可达的方式聚类,相比基于距离的聚类算法只能发现球状簇,该算法可以发现任意形状的聚类。

### 1.2 基于密度算法和 DSM 的模块聚类方法

聚类的目的是将众多对象通过一定的规律划分为不同的类别。针对于产品设计,则是通过 DSM 表达各个零部件间的关联关系,并以此为基础将零部件进行类别划分,每个类别便形成对应模块。由于参数选择的不同,会有不同的划分结果,此时再根据模块的划分准则进行判断,选出最优方案作为模块划分结果。

DSM 中的各个数值表示零部件间的关联关系强弱,数值大则表示关联关系强,反之则为弱。在聚类计算过程中,数值大则表示零部件间的距离紧凑,数值小则表示零部件间的距离松散。故以此为基础采用欧式距离计算两个零部件间的距离大小,其中欧式距离计算公式如下

$$d(x, y) = \sqrt{(x_1, y_1)^2 + (x_2, y_2)^2 + \dots + (x_n, y_n)^2} + \sqrt{\sum_{i=1}^n (x_i, y_i)^2} \quad (1)$$

式中:  $x_i$  为点  $x$  在各个维度中的距离;  $y_i$  为点  $y$  在各个维度中的距离。

若  $d(x, y) \leq \text{Eps}$ , 则表示两个零部件互相处于对方的 Eps 领域内。基于密度和 DSM 的算法步骤如下:

步骤 1, 输入设计结构矩阵, 并为每行数据进行编号, 用于识别零部件, 设置聚类参数, 计算各零部件间的欧氏距离;

步骤 2, 遍历 DSM 中所有零部件  $p$ , 如果该零部件  $p$  未被处理(未被归为某个簇), 则检查其 Eps 领域, 若其中包含的零部件数量大于 MinPts, 建立新簇  $C$ , 将 Eps 领域中所有的零部件加入簇  $C$  中;

步骤 3, 对簇  $C$  中其他零部件  $q$ , 检查其 Eps 领域, 若其中包含的零部件数量大于 MinPts, 则将其领域中未处理的零部件加入簇  $C$ ;

步骤 4, 重复步骤 3, 直到没有新的零部件加入当前簇  $C$  中;

步骤 5, 重复步骤 2 至 4, 直至遍历所有的零部件, 对遍历结束后未被处理的零部件依次建立并加入新簇  $C$ 。

针对 DSM 进行聚类, 算法中输入为矩阵形式且矩阵中的各个数值代表的是对应两个零部件间的关联关系。如表 1 所示为某产品的 DSM 矩阵,  $C_1, C_2, C_3, C_4, C_5, C_6$  表示该产品的 6 个零部件, 表中零部件的关联关系强弱由数值 0~1 进行量化表示。假设零件  $C_1$  对零件  $C_2$  存在影响关系, 表中 0.5 表示零件  $C_1$  对零件  $C_2$  影响关系的强弱程度大小, 若相比于零件  $C_2$ , 零件  $C_1$  对零件  $C_3$  的影响度更大, 则表中用 0.8 进行表示, 可以看出数值越大关联影响越强。所以表中 0 表示两个零件间没有关联影响, 1 表示零件自身对自身的影响度

最大。通过设计结构矩阵可以看出,零件间的关联影响不但是非布尔矩阵,同时关联影响也是非对称的(零件 A 对零件 B 的影响与零件 B 对零件 A 的影响不同)。故在聚类分析中对各个零部件间的距离判断不能直接采用矩阵中关联影响的数值。

由此利用  $n$  维空间的概念,将 DSM 矩阵中的数值转换为  $n$  维空间中的坐标,再求距离便可以解决以上问题。如表 1 所示共有 6 个零件,便可视作该矩阵空间共有 6 个维度,则零件  $C_1$  的坐标为  $(1,0.4,0.6,0,0,0)$ , $C_2$  的坐标为  $(0.5,1,0,0,0,0)$ ,其余同理。带入式(1)便可计算  $C_1$  和  $C_2$  间的欧式距离  $d(C_1,C_2)$  为

$$d(C_1,C_2)=\sqrt{\sum_{i=1}^n(x_i-y_i)^2}=\sqrt{(1-0.5)^2+(0.4-1)^2+(0.6-0)^2+(0-0)^2+(0-0)^2+(0-0)^2}\approx 0.9849$$

由于算法不需要输入也不识别零件号(即  $C_1,C_2,C_3,C_4,C_5,C_6$ ),故经过步骤 1 之后 DSM 矩阵变形为表 2 所示,第一列数字为零件的重新编号和识别。表 3 为各零部件间的欧氏距离计算结果。经过步骤 2 至 5 后,聚类结果如图 1 所示。

表 1 某产品 DSM 矩阵

Tab.1 Design structure matrix of a product

C	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	$C_6$
$C_1$	1	0.4	0.6	0	0	0
$C_2$	0.5	1	0	0	0	0
$C_3$	0.8	0	1	0	0	0
$C_4$	0	0	0	1	0	0.3
$C_5$	0	0	0	0.7	1	0.1
$C_6$	0	0	0	0	0.1	1

表 2 密度算法中的 DSM 矩阵

Tab.2 Design structure matrix of density algorithm

	1	1	0.4	0.6	0	0	0
2	0.5	1	0	0	0	0	0
3	0.8	0	1	0	0	0	0
4	0	0	0	1	0	0.3	
5	0	0	0	0.7	1	0.1	
6	0	0	0	0	0.1	1	

表 3 零部件间的欧氏距离计算结果

Tab.3 Euclidean distance between components

C	1	2	3	4	5	6
1	0.000 0	0.984 9	0.600 0	1.615 5	1.737 8	1.590 6
2	0.984 9	0.000 0	1.445 7	1.529 7	1.658 3	1.503 3
3	0.600 0	1.445 7	0.000 0	1.652 3	1.772 0	1.627 9
4	1.615 5	1.529 7	1.652 3	0.000 0	1.063 0	1.224 7
5	1.737 8	1.658 3	1.772 0	1.063 0	0.000 0	1.452 6
6	1.590 6	1.503 3	1.627 9	1.224 7	1.452 6	0.000 0

	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	$C_6$
$C_1$	1	0.4	0.6	0	0	0
$C_2$	0.5	1	0	0	0	0
$C_3$	0.8	0	1	0	0	0
$C_4$	0	0	0	1	0	0.3
$C_5$	0	0	0	0.7	1	0.1
$C_6$	0	0	0	0	0.1	1

图 1 聚类结果图

Fig.1 Clustering results

本次聚类设置  $MinPts=1, Eps=1.3$ , 所得聚类中心为零件 1 和零件 6。通过分析上述结果可以得出,当  $Eps$  为 1.3 时,零件 1 和零件 2,3 间距离小于  $Eps$ ,是直接密度可达,故聚为一类。零件 6 和零件 4 间距离小于  $Eps$ ,是直接密度可达,零件 4 和零件 5 间距离小于  $Eps$ ,是直接密度可达,则零件 6 和零件 5 是密度可达,故零件 6 和零件 4,5 聚为一类,至此所有零部件均被聚类处理。

### 1.3 模块的划分准则

通过调整聚类参数可以获得多种聚类结果,也就是获得多种模块划分方案,而方案的优劣则需要通过相关的评价指标进行判断。产品模块划分的驱动目标是为了使得模块内的零部件具有尽可能高的内聚度,而不同模块间零件的关联影响需要尽可能的小。故在此处采用文献[11]提出的一种模块度准则作为模块划

分准则,选择聚类结果中模块度最小的结果作为最终模块划分方案。

其提出的模块度计算方法是利用平均耦合度除以平均内聚度,模块度越小则说明模块划分方案更优。平均内聚度和平均耦合度的定义如下:

内聚度是指模块内部各个零部件的关联关系,假设某产品总共由  $m$  个模块组成,  $n_k$  和  $m_k$  分别为此产品第  $k$  个模块中的第一个零件和最后一个零件的序号,  $r(i, j)$  表示零件  $i$  和  $j$  的综合关联度,用  $I_k$  表示该模块的内聚度如下

$$I_k = \sum_{i=n_k}^{m_k} \sum_{j=n_k}^{m_k} r(i, j) / (m_k - n_k + 1)^2 \quad (2)$$

用  $I$  表示平均内聚度如下

$$I = \sum_{k=1}^m I_k = \left[ \sum_{k=1}^m \sum_{i=n_k}^{m_k} \sum_{j=n_k}^{m_k} r(i, j) / (m_k - n_k + 1)^2 \right] / m \quad (3)$$

耦合度是指两两模块之间的关联关系,假设  $n_p$  和  $m_p$  分别是第  $p$  个模块中第一个零件和最后一个零件的序号,则模块  $m_k$  和  $m_p$  之间的耦合度可用式(4)表示

$$E_{k,p} = \frac{\sum_{i=n_k}^{m_k} \sum_{j=n_p}^{m_p} r(i, j) + \sum_{i=n_p}^{m_p} \sum_{j=n_k}^{m_k} r(i, j)}{2(m_k - n_k + 1)(m_p - n_p + 1)} \quad (4)$$

式中:  $\sum_{i=n_k}^{m_k} \sum_{j=n_p}^{m_p} r(i, j)$  描绘了模块  $m_k$  对模块  $m_p$  的依赖度;  $\sum_{i=n_p}^{m_p} \sum_{j=n_k}^{m_k} r(i, j)$  描绘了模块  $m_k$  对模块  $m_p$  的影响度。

模块间的影响度和依赖度就构成两个模块间的耦合度。

式(4)可以确定产品任意两个模块之间的耦合度。产品的  $m$  个模块中有  $m(m-1)/2$  个两两模块,故所有模块间的平均耦合度为

$$E = \frac{\sum_{k=1}^{m-1} \sum_{p=k+1}^m E_{k,p}}{m(m-1)/2} \quad (5)$$

根据模块划分的基本原则,应使模块内聚度尽可能大,模块间耦合度尽可能小,即产品设计模块度最小。产品设计模块度体现了模块之间和模块内部的关联关系,从而可得模块度目标准则

$$F_M = E/I \quad (6)$$

## 2 应用实例

本文以某型汽车起重机上车部分为实例来说明模块划分过程,其主要零件如表4所示。

依照各零部件间的联接关联性、功能关联性和物理关联性3个方面,对汽车起重机上车部分各零件间关联关系进行DSM构建,所得结果如表5所示。

利用基于密度算法和DSM的模块聚类方法对表5所示的汽车起重机上车部分DSM矩阵进行模块聚类划分,参数设置:MinPts=1不变(表示一个模块至少包含一个零件),依次调整Eps=0.5560;Eps=0.4413;Eps=0.2954;Eps=0.2928;Eps=0.2817(Eps的大小可以依据各零部件间的欧式距离的计算结果进行相应选取)。根据以上参数选取,得到对应的模块聚类方案如表6所示。

以方案①为例,计算模块度数值  $F_M$ 。根据式(3)(5)(6)计算方案①的平均内聚度

$$I = \sum_{k=1}^m I_k = \left[ \sum_{k=1}^m \sum_{i=n_k}^{m_k} \sum_{j=n_k}^{m_k} r(i, j) / (m_k - n_k + 1)^2 \right] / m = 0.548329$$

平均耦合度

$$E = \frac{\sum_{k=1}^{m-1} \sum_{p=k+1}^m E_{k,p}}{m(m-1)/2} = 0.007779$$

表 4 汽车起重机上车部分主要零件  
Tab.4 Main components of truck crane

序号	名称	序号	名称
1	卷筒	20	变幅液压缸
2	卷筒轴	21	转台
3	卷筒离合器	22	回转轴承
4	卷筒制动器	23	滚珠
5	液压马达	24	人字架
6	联轴器	25	回转马达
7	液压马达减速器	26	回转马达制动器
8	吊钩	27	联轴器
9	吊钩横梁	28	回转马达减速器
10	拉板	29	驱动齿轮
11	滑轮组	30	起升液压缸
12	滑轮轴	31	起升控制阀
13	主臂	32	离合器控制阀
14	二节臂	33	起升背压平衡阀
15	三节臂	34	回转液压缸
16	伸缩液压缸	35	回转先导阀
17	缩回臂杆滑轮组	36	回转缓冲阀
18	伸出臂杆滑轮组	37	回转液动阀
19	定滑轮组		

表 5 汽车起重机上车部分 DSM 矩阵  
Tab.5 Design structure matrix of truck crane

C	1	2	3	4	5	6	7	...	34	35	36	37
1	1	0.33	0.20	0.20			0.13	...				
2	0.4	1	0.27			0.27		...				
3	0.4	0.27	1	0.13			0.13	...				
4	0.4		0.13	1				...				
5					1	0.13	0.07	...				
6		0.27			0.13	1	0.33	...				
7	0.33		0.13		0.27	0.13	1	...				
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
34								...	1	0.33	0.33	0.33
35								...	0.53	1	0.13	0.13
36								...	0.53	0.13	1	0.13
37								...	0.53	0.13	0.13	1

表6 模块聚类方案  
Tab.6 Module clustering scheme

方案	对应零件
①	{1,2,3,4,5,6,7},{8,9,10,11,12},{13,14,15,20},{16},{17,18},{19},{21,22,23,24},{25,26,27,28,29},{30,31,32,33},{34,35,36,37}
②	{1,2,3,4,6,7},{5},{8,9,10,11,12},{13,14,15,20},{16},{17,18},{19},{21,22,23,24},{25,26,27,28,29},{30,31,32,33},{34,35,36,37}
③	{1,2,3,4,6,7},{5},{8,9,10,11,12},{13,14,15,20},{16},{17,18},{19},{21,22,23},{24},{25,26,27,28,29},{30,31,32,33},{34,35,36,37}
④	{1,2,3,4},{5},{6,7},{8,9,10,11,12},{13,14,15,20},{16},{17,18},{19},{21,22,23},{10},{25,26,27,28,29},{30,31,32,33},{34,35,36,37}
⑤	{1,2,3,4},{5},{6},{7},{8,9,10,11,12},{13,14,15,20},{16},{17,18},{19},{21,22,23},{24},{25,26,27,28,29},{30,31,32,33},{34,35,36,37}

模块度

$$F_M = E/I = 0.014\ 187$$

同理可计算得到方案②至⑤的模块度  $F_M$ , 如表 7 所示。

表7 模块度  $F_M$

Tab.7 Modularity  $F_M$

参数	方案①	方案②	方案③	方案④	方案⑤
$I$	0.548 329	0.593 423	0.642 333	0.650 257	0.702 739
$E$	0.007 779	0.007 300	0.007 403	0.008 620	0.012 429
$F_M$	0.014 187	0.012 301	0.011 525	0.013 256	0.017 687

根据本文所提出的模块划分方法,应当选择 5 个方案中模块度数值最小的聚类方案作为本方法的最终模块划分结果。参考表 7 所得结果,其中方案③的模块度数值  $F_M=0.011\ 525$  最小,故选取方案③作为最终模块划分结果,即某型汽车起重机上车部分可以分为 12 个模块: $m_1=\{1,2,3,4,6,7\}$ , $m_2=\{5\}$ , $m_3=\{8,9,10,11,12\}$ , $m_4=\{13,14,15,20\}$ , $m_5=\{16\}$ , $m_6=\{17,18\}$ , $m_7=\{19\}$ , $m_8=\{21,22,23\}$ , $m_9=\{24\}$ , $m_{10}=\{25,26,27,28,29\}$ , $m_{11}=\{30,31,32,33\}$ , $m_{12}=\{34,35,36,37\}$ 。

### 3 结束语

模块划分是产品模块化设计中的核心部分,正确划分模块对产品性能、生产周期和成本有着显著影响。针对现有产品设计的实际情况,为满足高效快速获得模块划分方案的需求,本文提出了一种基于密度算法和 DSM 的模块划分方法。该方法可适用于所有的布尔、非布尔、对称、非对称矩阵,且计算速度快,参数设置简单,易于理解,并能保证所有零部件均参与聚类过程,同时引用了模块度准则作为评价指标,对聚类结果进行统一评价,获得最终模块划分方案。最后通过一个实例描述对所提出的模块划分方案进行验证分析,确定了该方案的正确性和可行性。

#### 参考文献:

- [1] 刘志,李帮义,程晋石,等. 基于模块化设计的制造/再制造生产决策[J]. 计算机集成制造系统,2016,22(4):935-944.
- [2] HÖLTTÄ OTTO K,CHIRIAC N A,LYSY D,et al. Comparative analysis of coupling modularity metrics[J]. Journal of Engineering Design,2012,23(10/11):790-806.
- [3] 谌炎辉,胡义华. 产品模块划分方法研究综述[J]. 机械设计与制造,2013(5):264-266.
- [4] 伍育红. 聚类算法综述[J]. 计算机科学,2015,42(S1):491-499.
- [5] ALGEDDAWY T,ELMARAGHY H. Optimum granularity level of modular product design architecture[J]. CIRP Annals-Manufacturing Technology,2013,62(1):151-154.
- [6] 刘建刚,唐敦兵,刘思峰,等. 基于联系信息流量的产品结构模块化方法[J]. 计算机集成制造系统,2011,17(11):2373-2382.
- [7] 马淑梅,黄立辉,李爱平,等. 运用改进模糊 C 均值算法的复杂产品功构单元模块划分方法[J]. 现代制造工程,2012(2):7-13.