

文章编号: 1674-8085(2017)04-0072-06

CBR 系统优化策略及在注塑模具设计中的应用

周敬勇¹, *谢世坤², 金曼曼²

(1.合肥职业技术学院, 安徽, 合肥 238000; 2.井冈山大学 机电工程学院, 江西, 吉安 343009)

摘 要: CBR (Case-based reasoning) 技术是弱理论强实践领域实现智能化的有效手段, 但是, 随着工业的飞速发展, 工程领域的特征信息爆炸式增长使得当前 CBR 系统的关键构建环节例如事例的表达、事例库的构建等呈现出诸多局限。为了适应新的复杂工程领域发展需要, 对若干关键环节进行了优化研究: 在 CBR 系统事例的表达方面, 利用 GHSOM (Growing Hierarchical Self-Organizing Map) 算法和粗糙集算法集成策略对事例庞大特征信息进行精简, 为事例知识的准确、高效表达、存储和检索打下基础; 利用 GHSOM 算法优化构建事例库, 通过空间聚类建构多层次多粒度知识库体系, 同时, 当执行事例检索时, 由 GHSOM 算法作为知识引导策略, 将新任务导入相应子事例库, 通过数值计算获取最佳事例。最后, 以注塑模具设计为应用领域验证了本方案的效果。

关键词: 粗糙集算法; 生长型分层自组织映射; 基于事例推理; 事例库; 注塑模具设计

中图分类号: TQ320.5, TP39

文献标识码: A

DOI:10.3969/j.issn.1674-8085.2017.04.013

OPTIMIZATION STRATEGY OF CBR SYSTEM AND ITS APPLICATION IN INJECTION MOLD DESIGN

ZHOU Jing-yong¹, *XIE Shi-kun², JIN Man-man²

(1. Hefei Technology College, Hefei, Anhui 238000, China; 2. School of Mechanical and Electrical Engineering, Jinggangshan University, Ji'an, Jiangxi 343009, China)

Abstract: CBR technology is an effective means to realize the intelligence of those domains, weak in theory but strong in practice. However, with rapid development in industry, characteristic information of engineering domain shows explosive growth, which makes the key links of CBR system building, such as case expression, case library construction and so on, are exposing many limitations. In order to meet the needs for the development of new complex engineering fields, this paper focuses on some key links for optimization: for case expression of CBR system, GHSOM algorithm and rough set algorithm are integrated to use in reducing the huge feature information of the case, which can lay a good foundation for correct and efficient expression, storage and retrieval of case knowledge. GHSOM algorithm is also used to optimize the construction of case library and by the spatial clustering of GHSOM, multi-level and multi-granularity of knowledge base system is constructed. At the same time, while case retrieval is executed, with GHSOM algorithm as a knowledge guidance strategy, new tasks are led into corresponding sub case library. Then, through numerical calculation the best examples are obtained. Finally, this paper takes the injection mold design as the application field to verify the effect of this scheme.

Key words: rough set algorithm; growing hierarchical self-organizing map; case-based reasoning; case base; injection mold design

收稿日期: 2017-05-16; 修改日期: 2017-06-15

基金项目: 国家自然科学基金项目(51165010); 江西省教育厅科技落地计划项目(KJLD14066); 江西省教育厅科技研究项目(GJJ160732)

作者简介: 周敬勇(1979-), 男, 河南信阳人, 讲师, 硕士, 主要从事模具 CAD/CAM/CAE 研究(E-mail: 171724260@qq.com);

*谢世坤(1973-), 男, 江西吉安人, 教授, 博士, 硕士生导师, 主要从事材料成型工艺控制的研究(E-mail: xskun@163.com);

金曼曼(1965-), 女, 江西峡江人, 副教授, 主要从事模具设计工作(E-mail: 604824496@qq.com).

0 引言

智能化是当今制造业发展的热点和高潮,世界各国相继提出本国的智能化发展策略和计划,例如德国的工业 4.0、我国的中国制造 2025、英国的工业 2050 战略和日本的科技工业联盟等。设计、制造等工程领域是极其复杂的强实践弱理论性领域,其知识难以表示成明确的规则或是模型^[1],使基于事例推理 (Case Based Reasoning, CBR) 技术成为其实现智能化的重要技术^[2]。

如何构建一个高质量的事例库,以及如何有效地检索最佳事例一直是构建 CBR 系统的关键,同时也是难点所在。目前有大量的研究聚焦在这两个方面,例如为了建立高效事例库,孙翎等^[3]通过粗糙集来对属性进行约简,从而简化事例表达特征集合,精简了事例信息量,不但促进了事例的存储和维护,而且大大提高了检索的效率。孙晋永^[4]等在定义 ALCQ (D) 范式规范事例的索引基础上,进行总分型相似度算法计算,从而可以更贴切地度量事例的相似性,提高事例检索的速度和准确性以及提高 CBR 系统的效率。其他还有很多研究例如文献[5-7]等对 CBR 系统的进步做出了有益的贡献。

但是,随着制造业的智能化、网络化、协同化发展,事例包含的原始特征信息爆炸式增长,仍用单一数据挖掘算法例如粗糙集等,不仅工作量巨大、效率很低,而且准确性急剧降低而无法完成特征精简任务。同时,在事例库的构建方面,目前的研究成果也不能适应新的情况。

为了使 CBR 技术适应新的形势要求,本文对相关环节进行了优化提升:在特征精简方面,本文将生长型分层自组织映射 (Growing Hierarchical Self-Organizing Map, GHSOM) 算法和粗糙集算法结合起来,首先利用 GHSOM 算法对特征进行聚类,然后对每一个子类利用粗糙集进行精简;在事例库构建方面,通过 GHSOM 空间聚类,快速、准确地形成事例库,并在检索时作为知识导引策略,使得事例的存储和检索统一起来,为实现系统应用的自动化、智能化打下基础。

1 基于 GHSOM 算法和粗糙集算法的事例特征的精简

1.1 GHSOM 算法的提出

SOM 即 Self Organizing Map,是一种优秀的无监督 ANN 聚类模型,是由学者 Kohonen^[8]提出的。SOM 的最超凡的能力是将一组高维数据根据其内在属性的差异进行聚类,在保持数据内在拓扑关系不变的情况下,映射到二维输出空间上,如图 1 所示, X 为投射在单位神经元上的 n 维输入向量, W 为相应的权重向量。

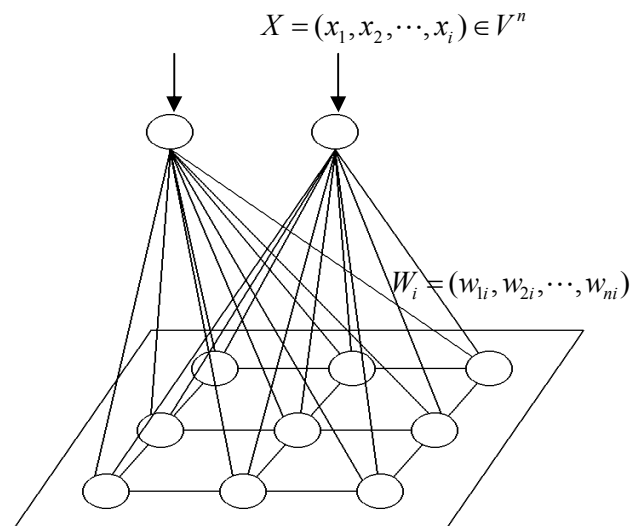


图 1 SOM 工作模型

Fig.1 SOM working model

然而, SOM 存在严重缺陷,那就是应用时必须首先知道起初数据的维数,如果不知道其维数,那么 SOM 聚类算法就无能为力;其次, SOM 聚类的结果仅限于二维平面性质的,对于每一个分类的结果,不能根据其内在性质的不同细化分成子类,即不具备空间分类属性。SOM 聚类的这些弱点给应用带来很多不便,因为在很多情况下,起初对数据的维数并不确定,并且通常需要进行空间分类^[9]。

为了能自发探测输入数据内在的组别属性并分配各组,而且对于每个子类根据实际情况连续细化分类,直至达到设定的条件,在 SOM 算法的基础上提出了 GHSOM,如图 2 所示。

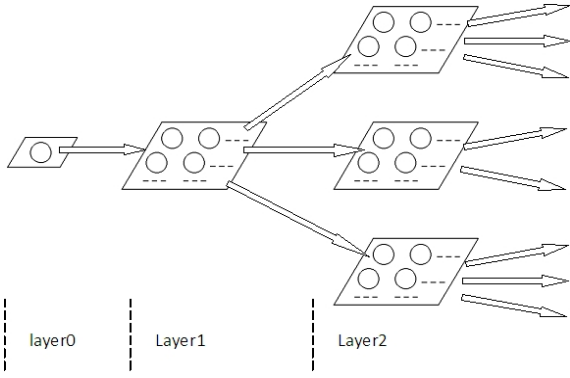


图2 GHSOM 工作模型
Fig.2 GHSOM working model

1.2 集成 GHSOM 算法和粗糙集算法的 CBR 系统事例特征精简策略

对于一个 CBR 系统来说，事例的表达是基础环节，而工程任务越来越繁多且复杂，其事例包含越来越庞大的信息，将这些信息全部表达并存入数据库是不现实的，并且会导致效率极低。因此如何从这些繁冗多杂的信息中分析出最本质的特征，用最小的本质属性特征集来全面、准确地表达 CBR 事例是很关键的。

粗糙集算法通过科学的计算属性约简，通过事物繁杂的现象属性找出其内在的本质属性，在不降低信息系统分类能力的基础上，能区分所有对象的最小属性子集代替原来的属性集，从而，成为 CBR 事例属性约简的广为应用的有效手段。

但是，随着目前工程任务包含的特征信息呈爆炸式增长，如果单纯用粗糙集约简，往往会因为计算量过大，造成效率和准确性大大降低。

为了解决这个问题，使 CBR 系统适应新的形势需求，本文将 GHSOM 算法和粗糙集算法集成起来实现事例特征集的约简，首先用 GHSOM 算法将庞大信息分为不相干的子类，然后对每一子类通过粗糙集算法进行约简。

1) 连续数据的离散化描述

设 $S=(U,A,V,F)$ 是一个信息系统， $A=C \cup D$ ， $C \cap D = \Phi$ ， C 称为条件属性集合， D 称为决策属性集。对于任意的一个属性 $a \in A$ ，定义一个序列 $v_1^a < v_2^a < \dots < v_k^a$ ，其中 $\{v_1^a, v_2^a, \dots, v_k^a\} = \{a(x); x \in U\}$ 。

对于决策信息系统 $S=(U,A,V,F)$ ， $A=C \cup D$ ， $a \in A$ ，令 $C_a = \{(a, c_1^a), (a, c_2^a), \dots, (a, c_k^a)\}$ ，为任意的

关于属性 a 离散区间。假定 $c_1^a < c_2^a < \dots < c_k^a$ ，离散断点 $C = \bigcup_{a \in A} C_a$ 定义了一个新的信息决策系统 $S^c = (U, A^c, d)$ ，称为 S 的 C 离散化，其中：

$$A^c = \{a^c; a \in A\}; \quad (1)$$

$$a^c(x) = \begin{cases} 0, & \text{当且仅当 } a(x) < c_1^a; \\ i, & \text{当且仅当 } a(x) \in [c_i^a; c_{i+1}^a], i \in \{1, k-1\} \\ k+1, & \text{当且仅当 } a(x) < c_k^a; \end{cases} \quad (2)$$

2) 定性特征的定量化表达

在所有的特征中，不是所有的特征都可以定量计算，部分特征是定性描述，即通过专家评价获得评语，作为模糊语言值。评语集是对评价对象做出的可能评价结果所组成的集合，表示为 $V = \{v_1, v_2, v_3, v_4, v_5\}$ 。考虑评价等级的合理密度，在特征指标体系中采用五级评语集，即 $V = \{\text{很好, 较好, 中, 较差, 很差}\}$ 。

对于特征的各个评价因素，系统给出具体量化。当所有的定性指标给出定性评价后，需要将定性评价量化，并建立隶属度函数。定性评价量化赋值按表 1 进行。

表 1 定性评价量化赋值

模型性能指标评语	很好	较好	中	较差	很差
对应数值	9	7	5	3	1

相应的隶属度函数式如式 (3) 所示：

$$r_{ij} = \begin{cases} 1, & x_{ij} \leq a; \\ \frac{b - x_{ij}}{b - a}, & a < x_{ij} < b, i = 1, 2, k, m, \quad j = 1, 2, k, n; \\ 0, & x_{ij} \geq b; \end{cases} \quad (3)$$

3) GHSOM 算法过程参数设置

GHSOM 算法在涉及到几个重要的参数，根据不同的情况进行相应设置：

① mqe_i ：神经元 i 的权重向量和投射在该神经元上的向量的误差的算术平均值，其计算公式如式 (4) 所示：

$$MQE_m = \frac{1}{n} \sum_{i \in U} mqe_i \quad (4)$$

u 表示第 m 层的神经元分组数, n 表示该层所有的神经元总数目。如果 mqe_i 的值较大, 则表示投射到该单元 i 的输入向量不很精确, 需要加入新的单元。

② mqe_0 : 第 0 层单一神经元的算术平均误差, 其计算公式如式 (5) 所示:

$$mqe_0 = \frac{1}{n_l} \sum_{x_j \in l} \|w_0 - x_j\| \quad (5)$$

其中, n_l 是所有输入向量的数目。

③ MQE (Mean quantization error): 第 m 层所有神经元的算术平均误差的均值, 其计算公式如式 (6) 所示:

$$mqe_i = \frac{1}{n_{D_i}} \sum_{x_j \in D_i} \|w_i - x_j\| \quad (6)$$

其中, n_{D_i} 是第 i 层神经单元的数目。

④ 最不相似邻居 d : 在最粗糙单元 (最大误差单元) e 的所有相邻神经元中, 其权重向量与最大误差单元的权重向量距离最大的单元。

4) 基于 GHSOM 和粗糙集的事例特征提取算法过程

首先将事例的所有信息通过 GHSOM 进行聚类, 获得若干子类。对于每一个子类, 根据属性重要程度从小到大进行排序; 在属性相同的情况下, 按照条件属性断点个数由多到少进行条件属性的排序; 其次, 对于属性中的每个断点进行条件判断: 把信息系统中与当前断点相邻的两个属性值进行排序, 然后将小的值改成较大的数值, 如果信息系统不冲突, 则消除此断点, 否则把修改过的属性值还原; 第三, 对每个属性进行步骤 2 的循环迭代: 算法对每个断点进行判定, 去掉冗余的断点, 得到离散的信息系统。

2 基于 GHSOM 的事例库建构

当原始案例数据经过特征提取、数据转换及标准化后, 作为 GHSOM 系统的输入空间向量通过聚类进行事例库的建构, 具体步骤如下:

Step 1: 初始化 GHSOM 系统的参数, 包括学

习率, 临近范围、训练过程的起始层大小以及停止标准等;

Step 2: 建立第 0 层。该层实际上是由一个神经元构成的虚拟层, 其权重初始化为输入数据对应权重的平均值。利用欧几里得公式计算该单元权重向量与输入单元权重向量的 mqe (the mean quantization error, 算术平均误差);

Step 3: 建立第 1 层。首先设立一个 2×2 维大小的初始层, 并按照 SOM 训练过程进行神经网络训练。计算该层上每一个神经元的权重向量与映射到该单元的输入单元的权重向量的误差算术平均值, 确定具有最大值的单元即为误差最大单元。

比较误差最大单元与其相邻单元的相似度, 差别最大的即为最不相似单元。于是在该单元与最不相似单元之间插入新的单元, 每一次加入新的单元, 都要计算该层所有单元的权重向量误差算术平均值 MQE (Mean quantization error), 重复进行直至该层的 MQE 值降到该单元的 mqe 值的某一分数 α 为止, 即满足公式 (7)。该层的 MQE 值越低, 则该层被训练得越好。

$$MQE_m < \alpha gmqe_u \quad (7)$$

Step 4: 建立后续层。对当前层的每一个单元进行映射, 过程和 Step 3 一样, 直到当前层 (设为第 i 层) 的 mqe (mqe_i) 小于第 0 层的 mqe (mqe_0) 的某一分数 β , 如公式 (8) 所示, 层次扩展停止。

$$MQE_m < \beta gmqe_u \quad (8)$$

Step 5: 通过重复执行前面 4 步, 初始事例被分配到相应的子事例库中。通过改变 α 和 β 的值, 可以改变子事例库的分类粒度。显然, 子事例库数目越多, 同一子事例库中的事例差别就越小, 子事例库就越精致。一般情况下, 通过 GHSOM 系统将整个事例库分为三层就可以了, 但如果事例过于复杂和庞大, 可以多分几层, 但通常不超过 5 层。

3 基于 GHSOM 的相似事例检索机理

当一个新的设计任务产生时, 它首先被转换成事例向量模型, 然后进行数据标准化。这样, 就由

新问题形成一个新的输入向量, 经由 GHSOM 系统导入相应的子事例库。在该子事例库中, 通过最邻近测量方法中的欧几里得算法计算新事例与子事例库的各事例的数值相似度, 获取最佳事例。欧几里得算法用向量余弦公式计算^[10], 表示如下:

假设向量 $X=[x_1, x_2, \dots, x_i, \dots, x_n]$ 和 $Y=[y_1, y_2, \dots, y_i, \dots, y_n]$, 那么二者的相似度计算如公式 (9) 所示:

$$Sim(X, Y) = \cos ine(X, Y) = \frac{X \cdot Y}{\|X\| \|Y\|} \quad (9)$$

综上所述, 在本文中, CBR 系统的检索由两级构成, 第一级即 GHSOM 系统的导入策略, 作用类似于传统 CBR 系统的知识导引法, 只是其建立在科学算法基础上, 更精确、高效。通过第一级检索, 新事例与相应子事例库的事例相似性较高, 大大缩减了第二级检索的搜索空间, 从而使检索效率和质量都显著提高。

4 系统的实现及应用实例

随着塑料性能的不断提高, 注塑产品加速成为诸多产品的替代品, 高质量、高效率地生产塑料产品使得注塑模具^[11-12]智能化设计成为当务之急。由于注塑模具设计是一个弱理论强实践领域, 因此, 独具优势的CBR技术成为实现该领域智能化的最佳选择。

在前面的基本理论、基本方法研究的基础上, 利用 Microsoft SQL Server 2000 中 Analysis Manager 组件的开放接口^[13]以及Jana、VC等编程工具, 采用模块化开发策略, 开发了集成GHSOM和粗糙集的注塑模具设计原型系统。系统共分为三个模块: 注塑模具设计任务分解模块, 其功能主要是将设计任务分解成相互独立的子任务, 每一个子任务对应一个相对独立的结构, 并负责将设计任务转化为适合数据挖掘的形式; 数据挖掘模块, 其作用主要是负责获取每一个子任务对应的模具结构; 模具成形模块, 其负责将每一个子结构修改为恰当的形式, 并使各子结构相互协调形成注塑模具总体结构。用户界面如图3所示。

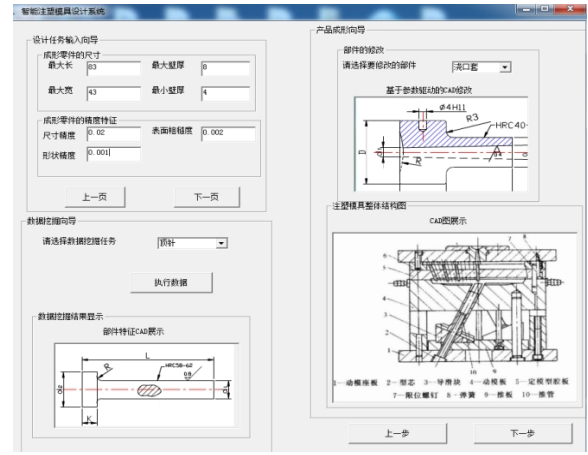


图3 系统用户界面图

Fig.3 System user interface

为了客观、准确地评价所构建系统的实际性能与效果, 选用多个典型注塑模具设计任务作为应用实例进行实验验证, 这些设计任务在实际生产中很常见, 有大量现存案例以及相关数据记录, 便于进行指标对比。如图3所示, 运行的为其中一个案例——塑料盖的注塑模具设计, 其他还有塑料防护罩、手机壳、塑胶碗等注塑模具设计^[14]。评价指标选用准确率(利用置信区间表示)、效率(所耗费的时间)和耗费成本(金额的区间值)。这些指标值是以统计的方式, 通过统计多项设计任务的相应数据, 进行加和求平均的方法获取, 结果如表2所示^[15-16]。

表2 本文建议方法与现存典型方法的效果比较

Table 2 Effect comparison of the proposed method in this paper and the existing typical method

评价指标平均值 应用的设计方法	准确率置信 区间	耗费时 间(H)	耗费成本 (万)
专家人工设计	[0.70, 0.90]	20	0.7-1.5
传统的智能设计方法	[0.75, 0.86]	2	0.5-0.9
本文建议的方法	[0.90, 1]	0.5	0.05-0.10

从表2可以明显看出, 本文建议的系统每次成功的概率远远高于传统的方法, 更为突出的是其效率和成本远远低于传统的方法。因此, 本文建议的方法具有重要的应用价值, 并且它的性能还会随着应用不断提高。另外, 本方法具有普遍意义, 可以广泛应用于其他弱理论强实践性领域。

4 结论

1) 特征提取准确与否决定着 CBR 系统的事例能否准确表达以及后续的有效检索, 是 CBR 系统

的关键环节之一,利用粗糙集和 GHSOM 等数据挖掘算法对 CBR 事例进行冗余特征去除,克服了传统单一使用粗糙集进行特征约见难以应用大量特征信息,从而使 CBR 系统能适应日益复杂的工程领域的发展,为系统成功实施提供了保障。

2) 利用 GHSOM 挖掘算法构建 CBR 系统事例库,一方面充分发挥 GHSOM 算法能根据输入数据的内在拓扑结构自动进行多层次多粒度知识的分类;另一方面,在知识检索时它也能起知识导引作用。从而,通过 GHSOM 算法的应用使 CBR 系统知识的存储和检索无缝结合起来,克服了传统系统因为知识导引和知识存储的不兼容、不连贯而造成的检索效率及准确性降低的缺陷。

3) 利用数据挖掘技术改造优化了 CBR 系统结构的特征提取、事例检索等环节,取得显著的效果。但是,在利用数据挖掘技术对事例进行修改优化方面的研究尚不充分,下一步的研究将重在利用数据挖掘中的遗传算法对事例进行修改优化。

参考文献:

- [1] Ding H, Yang Z. Method and system of shearer cutting unit modern design oriented to KBE[J]. Meitan Xuebao/journal of the China Coal Society, 2012, 37(10):1765-1770.
- [2] PING H. Collaborative case-based reasoning for knowledge discovery of elders health assessment system[J]. Open Biomedical Engineering Journal, 2014, 8(1):2311-2319.
- [3] 孙翎,张金隆,迟嘉昱. 基于粗糙集的 CBR 系统案例特征项权值确定[J]. 计算机工程与应用,2003,39(30): 44-46.
- [4] 孙晋永,古天龙,常亮,等. 基于 ALCQ(D)的 CBR 事例表示及相似性度量[J]. 计算机科学, 2014, 41(4):223-229.
- [5] 万磊,孙继银,王莎莎. 基于粗糙集的 CBR 系统案例特征属性优化算法研究[J]. 微电子学与计算机,2008, 25(1):190-192.
- [6] Abdullah A B. Profile Measurement Based on Focus Variation Method for Geometrical Defect Evaluation a Case study of Cold Forged Propeller Blade[J]. Advances in Mechanical Engineering, 2014, 68(1): 174-183.
- [7] 季赛,沈星,沈超. 基于粗糙集和相似度量的 CBR 检索方法[J]. 计算机工程与应用, 2006, 42(13):172-174.
- [8] Kohonen T. Self-organized formation of topologically correct feature maps. Biological Cybernetics, 1982, 43(2): 1241-1251.
- [9] Guo Y, HU J, Peng Y. Research on CBR system based on data mining [J]. Applied Soft Computing, 2011, 11(8): 5006-5014.
- [10] Salton G, Mcgill M J. Introduction to Modern Information Retrieval. Auckland: McGraw-Hill, 1983, 13(4): 2134-2142.
- [11] 谢世坤,郭秀艳,易荣喜,等. 基于UG的遥控器上下壳体注塑模设计[J]. 井冈山大学学报:自然科学版,2015,36(1):61-64.
- [12] 郭秀艳,谢世坤,高芝,等. 基于 CAD 的透盖注塑模具设计[J]. 井冈山大学学报:自然科学版,2014, 35(5):75-78.
- [13] Wu k, Liu F Y. Application of Data Mining in Customer Relationship Management [J]. Journal of Ningbo olytechnic, 2008, 12(3): 1321-1328.
- [14] 王成玥,陈振,黄勇,等. 基于Moldflow的底座注塑成型数值模拟及工艺优化[J]. 塑料工业,2016,44(10): 34-38.
- [15] 贾颖莲,何世松. Pro/E在注塑模具设计中的研究与应用[J]. 煤矿机械, 2007, 28(5): 75-77.
- [16] Chao L, Yan Y, Feng L. Research and Development of Die-casting Die CAD System Based on CBR[J]. Hot Working Technology, 2010, 39(5):60-62.