

Research on Optimization of Design Domain based on Decision Tree and its Application to Vehicle Optimization Design

Hu Chen¹, Zhan Zhenfei¹, Yang Junqi¹, Li Jie², Yu Huili²

¹College of Automotive Engineering, Chongqing University, Chongqing, China, 400044

²Chongqing Changan Automobile Co., Ltd, Chongqing, China, 400023

Email:zhenfeizhan@cqu.edu.cn

Abstract: Design of Vehicle weight reduction has a lot of features such as: large amounts of design variables, complex system structure and nonlinear response. The traditional optimal algorithms used in optimal design easily lead to huge calculation, poor efficiency and even no access to get an optimal solution. To solve these problems, this paper proposes a method based on decision tree to optimize design domain. By this method, the original design domain is classified to facilitate the designers. It not only reduces the ranges of the initial design points in the optimal algorithm but also improves the design efficiency of the optimal algorithm.

Keywords: decision tree; design domain; reduced design domain; convergence efficiency

基于决策树的设计空间优化研究及其在汽车优化设计中的运用

胡宸¹, 詹振飞¹, 杨俊祺¹, 李洁², 禹慧丽²

¹重庆大学汽车工程学院, 重庆, 中国, 400044

²重庆长安汽车股份有限公司, 重庆, 中国, 400023

Email: zhenfeizhan@cqu.edu.cn

摘要: 目前汽车轻量化设计具有设计变量多、系统结构复杂以及响应非线性强等特点, 利用各类优化算法求解, 计算量大且效率低, 甚至难以获取有效优化解。针对上述问题, 本文提出基于决策树数据挖掘技术的设计空间优化方法, 对原始设计域进行划分, 缩小优化算法对初始设计点的起始范围, 提高了优化算法的设计效率。

关键词: 决策树; 设计空间; 缩减域; 收敛效率

1 引言

汽车车身结构轻量化设计往往具有设计变量多约束条件多等特点, 其优化数学模型形态较为复杂。针对这样的优化问题, 各类优化算法对初始设计点的依赖程度较高, 初始点选取的好坏决定了优化算法最终收敛与否, 其收敛效率直接取决于初始设计点的选取位置。有效地设计空间对优化算法的收敛性有促进作用, 科学合理地布置初始设计点, 可以有效的减少寻找优化解的计算量, 节约时间和设计资源, 从而提高优化算法寻找有效优化解的效率。

目前有很多研究者试图通过各种方法来得到可行的缩减域, 从而提高优化算法过程中的效率, BATES 等^[1]利用克里格方法来构造可行设计域, 并指出可行域从范围形状上来讲可能并不规则, 甚至是由几个不连续的空间所组成。K.Wang 等^[2]指出传统的遗传算法对于高约束问题并不适用, 提出利用数据分析得到关于可行域的更多信息并基于此定义出明确的交叉变异后, 得到的定向的遗传算法可以有效地解决这一问题。

现阶段针对可行域的研究大多是利用数据挖掘技术来鉴别原始设计域的特征以得到可行域从而来提高优化算法过程中的计算效率。常用的数据挖掘技术包括了高斯过程、支持向量机以及决策树等。Kuss 等^[3]利用高斯过程来定义了灵活有效的分类模型。Basudhar 等^[4]利用支持向量机定义了一种极限状态方程的结构, 其可以有效改善系统

在整个优化过程中的表现，保证优化设计的稳健性和可靠性。Breiman 等^[5]介绍了使用分类与回归树法来构建决策树模型，进而对响应变量进行预测。诚然，各种方法都有各自的特点和优势，决策树和前两者的最大区别就在于它是一种非参数法，研究者可以利用它针对特定的目标得到一些特定的模式，基于这样的特点，决策法已经在很多领域中都得到了充分利用。

本文针对汽车车身优化设计，基于适当的试验设计方法对设计变量的范围进行划分，利用决策树对各个划分后的设计空间的合理性和可行性进行探究并最终得到缩减可行域，然后进一步验证所得缩减可行域在寻找优化解的过程中是否对优化算法找到收敛解并提高设计效率具有促进作用。

2 分类与回归树分类方法

分类与回归树是 Breiman 等^[5]人于 1984 年提出的一种决策树分类方法。其选择具有最小基尼指数(Gini Index)值的属性作为测试属性，从而产生二叉树。一般的，利用决策树来分类主要分为两个阶段，第一阶段需要利用训练样本集构建决策树模型，该过程是一个从数据中获取知识并进行机器学习的过程；第二阶段就是基于建立好的决策树模型对未知数据样本集进行分类，根据根节点对该对象的属性逐渐测试其值，顺着分支向下走直到达到某一个叶节点，此叶节点所代表的类别就是该对象所处的类别。

分类与回归树方法的思想就是通过节点的不纯度指标来对不同类别的数据进行分类的，在分类与回归树中，我们利用 Gini 来度量节点的不纯度，其中一个节点的不纯度定义为：

$$i(t) = \sum_{k \neq l} p(k|t)p(l|t) = 1 - \sum_{k=l} (p(k|t))^2 \quad (1)$$

其中， k, l 为类别指标， $p(k|t)$ 和 $p(l|t)$ 为数据中不同类别数据在节点 k 处出现的条件概率。

基于 Gini 分裂准则，求解左右子节点不纯度变差，对于不纯度的降低定义为：

$$\Delta i(s, t) = i(t) - p_L i(t_L) - p_R i(t_R) \quad (2)$$

其中， p_L 和 p_R 是一个样本出现在左子节点 t_L 和右子节点 t_R 的概率。

分类回归树使用了自顶而下的递归方式，在决策树的内部节点进行属性值的比较和判断，并依据属性值确定该节点向下的分支。本文采用决策树数据挖掘技术来得到缩减可行域的基本思路是：构建决策树模型时将现有的目标函数值根据其大小来进行分类，本文中按照目标函数值从小到大排列后，并根据所研究问题的实际需求，将它们分为了三个类别，以字母“G”、“F”和“B”来表示，分别代表好、一般和差三种类型。分类完成以后，便利用这些训练样本集来构建决策树模型；决策树模型构建完成后，就需要进行原始设计空间的划分以及各缩减域的验证工作，以保证最终能够得到可行缩减域。本文中将原始设计域进行四等分，进而通过适当的试验设计方法在每个划分后的设计空间里产生一定数量的随机样本，计算出它们对应的目标函数值。再利用建立好的决策树模型来对得到的所有目标函数值进行分类，对每个划分后的设计空间进行验证和评估，从而最终得到所需的可行缩减域。

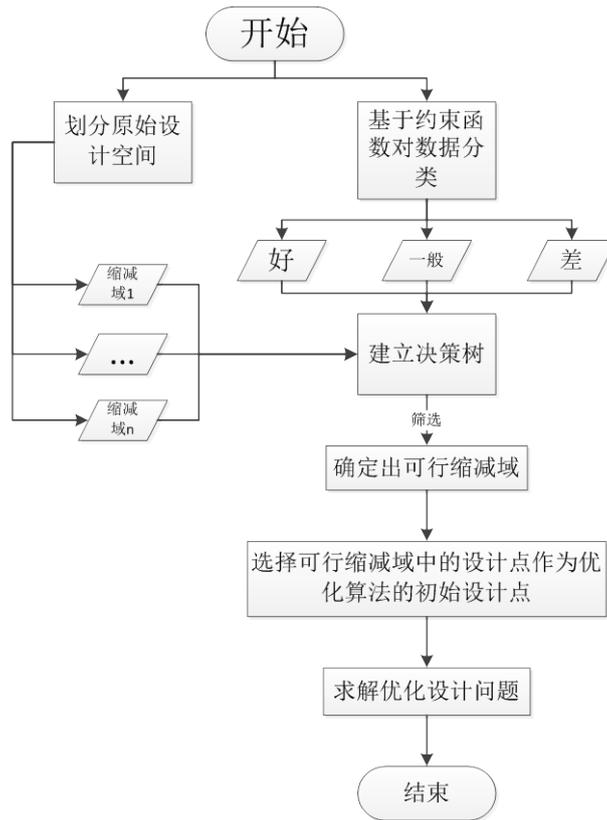


Figure1. Flowchart of design domain based on decision trees
图 1. 基于决策树方法的设计域优化流程

3 工程实例

基于设计空间优化的汽车优化设计旨在利用得到的缩减可行域，减少优化过程中得到有效优化的迭代次数，从而提高优化算法的优化效率以及获取有效优化的能力。本节利用 2001 Ford Taurus 的有限元模型碰撞实验中的 B 柱加速度和碰撞有效距离两项实验数据来开展本文的研究内容。根据实际要求，B 柱加速度值的绝对值应小于 38，有效碰撞距离在合理范围内越小越好。

基于决策树数据挖掘技术的设计空间优化的主要实现步骤为：

(1) 对所得实验数据进行分类，并分别建立对应的决策树；

(2) 划分原始设计空间：将原始设计空间中的 8 个变量都均分为 4 个小范围，通过适当的试验设计方法，分别从 8 个变量各自的 4 个小范围中选取一个，得到一系列新的设计空间。

(3) 缩减域的评估：得到一系列新的设计空间后，通过适当的试验设计方法，在每个缩减域中生成 10 个样本点，并利用响应面模型得到它们的目标函数值，得到 B 柱加速度和有效碰撞距离两个响应变量的全部函数值，再利用建立好的决策树对全部目标值分类，根据分类结果进一步判断各个缩减域的优劣及可行性。

(4) 分别选择原始设计空间和缩减设计域中的设计点作为优化算法中的初始设计点，比较两者得到有效优化的设计效率；

(5) 采用 NSGA-II 算法进行求解：遗传算法是一种仿生进化算法，针对优化问题，使用遗传算法求解是一个非常好的选择。遗传算法的一般步骤为编码、种群初始化、适应度计算、选择、交叉、变异和解码等。NSGA-II 算法是针对 NSGA 算法的缺陷提出的改良算法，NSGA-II 提出了快速非支配排序方法，降低了算法的复杂度；采用精英策略，扩大采样空间，使父代和子代种群组合，共同参与竞争产生下一代种群，容易得到更加优良的下一代。

3.1 建立决策树

根据本文的实验数据以及研究的输出响应量 B 柱加速度和有效碰撞距离的实际约束, 将数据总共分为三类, 分别用数字“1”“2”和“3”来表示, 分类完成后建立决策树。

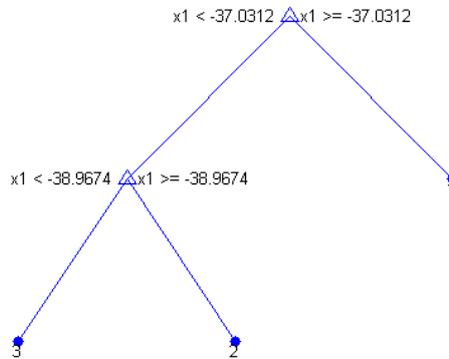


Figure 2. Decision tree of B pillar acceleration
图 2. B 柱加速度决策树

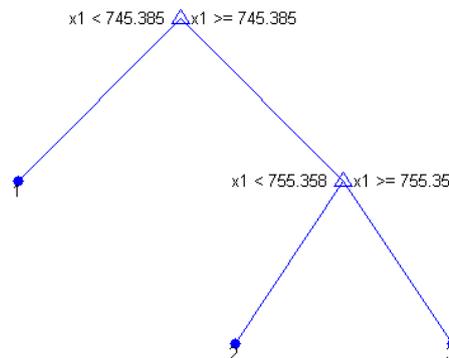


Figure3. Decision tree of crash distance
图 3. 有效碰撞距离决策树

3.2 划分原始设计空间

本文的实验数据中, 每一个样本有 8 个输入变量, 2 个输出响应量, 为了提高寻找优化解过程中的设计效率, 减少寻优过程中的迭代次数, 我们将 8 个变量的原始范围都均分为 4 个小范围, 然后基于适当的试验设计方法, 分别从 $x_1, x_2, x_3, x_4, \dots, x_8$ 中各自的四个小范围中选取一个, 组合形成一个新的设计空间。

其中, 原始设计空间为:

如表 1 所示为原始设计空间的范围, 进而利用适当的试验设计方法对原始设计空间进行划分, 得到 15 组子设计空间, 所得各缩减域如下表所示:

Table1. Original design domain

表 1. 原始设计空间

	x1	x2	x3	x4	x5	x6	x7	x8
Lower Bound	1.5	1.5	2.0	1.9	2.0	1.8	1.8	1.2
Upper Bound	2.3	2.3	3.0	2.9	3.1	2.7	2.7	1.8

Table2. Reduced design domain

表 2. 缩减设计空间

x1	x2	x3	x4	x5	x6	x7	x8
(1.5,1.7)	(1.7,1.9)	(2.5,2.75)	(2.65,2.9)	(2.0,2.275)	(2.475,2.7)	(2.475,2.7)	(1.5,1.65)
(1.7,1.9)	(1.5,1.7)	(2.25,2.5)	(1.9,2.15)	(2.275,2.55)	(1.8,2.025)	(2.025,2.25)	(1.5,1.65)
(1.9,2.1)	(1.7,1.9)	(2.0,2.25)	(2.65,2.9)	(2.55,2.825)	(2.25,2.475)	(2.25,2.475)	(1.2,1.35)
(2.1,2.3)	(1.7,1.9)	(2.75,3.0)	(2.15,2.4)	(2.275,2.55)	(2.475,2.7)	(2.025,2.25)	(1.35,1.5)
(2.1,2.3)	(1.7,1.9)	(2.0,2.25)	(2.15,2.4)	(2.55,2.855)	(2.025,2.25)	(2.25,2.475)	(1.2,1.35)
(1.7,1.9)	(2.1,2.3)	(2.5,2.75)	(1.9,2.15)	(2.825,3.1)	(2.475,2.7)	(2.25,2.475)	(1.65,1.8)
(1.9,2.1)	(2.1,2.3)	(2.0,2.25)	(2.4,2.65)	(2.0,2.275)	(1.8,2.025)	(2.475,2.7)	(1.35,1.5)
(1.9,2.1)	(1.9,2.1)	(2.5,2.75)	(2.15,2.4)	(2.55,2.825)	(1.8,2.025)	(2.475,2.7)	(1.5,1.65)
(1.7,1.9)	(1.9,2.1)	(2.75,3.0)	(1.9,2.15)	(2.55,2.825)	(2.475,2.7)	(2.25,2.475)	(1.5,1.65)
(1.9,2.1)	(1.9,2.1)	(2.25,2.5)	(1.9,2.15)	(2.275,2.55)	(2.25,2.475)	(1.8,2.025)	(1.35,1.5)
(2.1,2.3)	(2.1,2.3)	(2.0,2.25)	(2.15,2.4)	(2.825,3.1)	(2.25,2.475)	(2.475,2.7)	(1.35,1.5)
(1.5,1.7)	(1.9,2.1)	(2.5,2.75)	(2.4,2.65)	(2.0,2.275)	(1.8,2.025)	(1.8,2.025)	(1.65,1.8)
(1.5,1.7)	(2.1,2.3)	(2.75,3.0)	(2.4,2.65)	(2.825,3.1)	(2.25,2.475)	(1.8,2.025)	(1.65,1.8)
(1.5,1.7)	(1.5,1.7)	(2.25,2.5)	(2.65,2.9)	(2.0,2.275)	(2.025,2.25)	(2.025,2.25)	(1.2,1.35)
(1.9,2.1)	(1.5,1.7)	(2.75,3.0)	(2.65,2.9)	(2.825,3.1)	(2.025,2.25)	(1.8,2.025)	(1.35,1.5)

3.3 评估缩减域

通过适当的试验设计方法得到 15 个缩减设计空间后，需要对各个新的设计域进行评估，进一步验证它们的优劣及可行性，获得最终的缩减可行域。

在每一个缩减设计空间中利用适当的试验设计方法生成 10 个样本点，并得到这 10 个样本的目标函数值。改变设计变量的范围，重复上述步骤获得全部缩减设计域中 B 柱加速度和碰撞有效距离两个响应变量的全部函数值，进而利用已经建好的决策树对所得数据进行分类，根据两个响应变量对应的分类结果，判别 15 个缩减设计空间的优劣，从而最终得到缩减可行域。B 柱加速度数据分类结果和有效碰撞距离数据分类结果分别如表 3 和表 4 所示：

Table3. Results of B pillar acceleration data classification
表 3. B 柱加速度数据分类结果

<i>GR1</i>	<i>GR2</i>	<i>GR3</i>	<i>GR4</i>	<i>GR5</i>	<i>GR6</i>	<i>GR7</i>	<i>GR8</i>	<i>GR9</i>	<i>GR10</i>	<i>GR11</i>	<i>GR12</i>	<i>GR13</i>	<i>GR14</i>	<i>GR15</i>
F	G	G	F	G	F	G	G	F	G	G	F	F	F	F
B	F	F	F	G	F	G	F	B	F	G	F	F	F	F
F	G	G	F	G	F	G	G	F	G	G	F	F	F	F
B	G	G	F	G	F	G	G	F	G	G	F	F	F	F
F	G	G	F	G	F	G	G	F	G	G	F	F	F	F
F	G	G	F	G	F	G	G	F	G	G	F	F	F	F
B	F	G	F	G	F	G	G	B	G	G	F	F	F	F
B	F	G	F	G	F	G	G	B	G	G	F	F	F	F
F	G	G	F	G	F	G	G	F	G	G	F	F	F	F
B	G	G	F	G	F	G	G	F	G	G	F	F	F	F

Table4. Results of crash distance data classification
表 4. 有效碰撞距离数据分类结果

<i>GR1</i>	<i>GR2</i>	<i>GR3</i>	<i>GR4</i>	<i>GR5</i>	<i>GR6</i>	<i>GR7</i>	<i>GR8</i>	<i>GR9</i>	<i>GR10</i>	<i>GR11</i>	<i>GR12</i>	<i>GR13</i>	<i>GR14</i>	<i>GR15</i>
F	B	G	G	G	G	G	G	G	G	G	B	G	B	F
F	B	G	G	G	G	G	G	G	G	G	B	G	B	F
F	B	G	G	G	G	G	G	G	G	G	B	G	B	F
F	B	F	G	G	G	F	F	G	G	G	B	F	B	F
F	B	G	G	G	G	G	F	G	G	G	B	F	B	F
G	B	G	G	G	G	G	G	G	G	G	F	G	B	G
F	B	G	G	G	G	G	G	G	G	G	B	G	B	F
F	B	G	G	G	G	G	G	G	G	G	B	G	B	G
F	B	G	G	G	G	G	G	G	G	G	B	G	B	F
F	B	G	G	G	G	G	G	G	G	G	B	G	B	F

其中 GR 表示组别，G、F 和 B 分表表示分类结果好、一般和差三种类型。

综合考虑 B 柱加速度和有效碰撞距离两个响应变量的数据的分类结果，获得缩减可行域，如表 5 所示。

Table5. The range of design variable in reduced design domain

表 5. 缩减设计空间设计变量范围

优化设计空间		ODD		
		R0	R1	R2
设计变量范围	x1	[1.50,2.30]	[2.10,2.30]	[2.10,2.30]
	x2	[1.50,2.30]	[1.70,1.90]	[2.10,2.30]
	x3	[2.00,3.00]	[2.00,2.25]	[2.00,2.25]
	x4	[1.90,2.90]	[2.15,2.40]	[2.15,2.40]
	x5	[2.00,3.10]	[2.55,2.82]	[2.82,3.10]
	x6	[1.80,2.70]	[2.02,2.25]	[2.25,2.47]
	x7	[1.80,2.70]	[2.25,2.47]	[2.47,2.70]
	x8	[1.20,1.80]	[1.20,1.35]	[1.35,1.50]

3.4 优化设计结果分析

根据 B 柱加速度和碰撞有效距离两个响应量的实际约束条件，采用 NSGA-II 算法来求解有效优化解，并分别选择原始设计空间和缩减可行域中的设计点作为优化算法中的初始设计点，并比较两者得到有效优化解的设计效率。

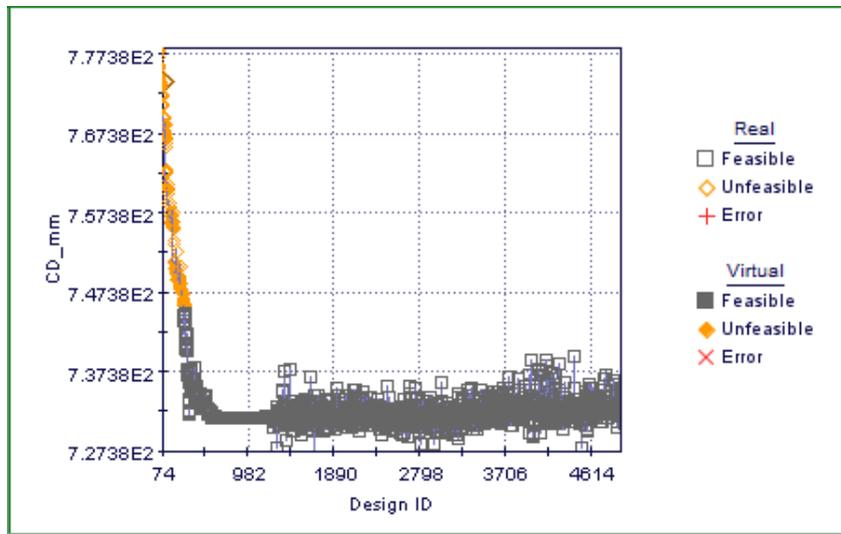


Figure4. Optimization design status based on original design domain

图 4. 基于原始域的优化设计状态

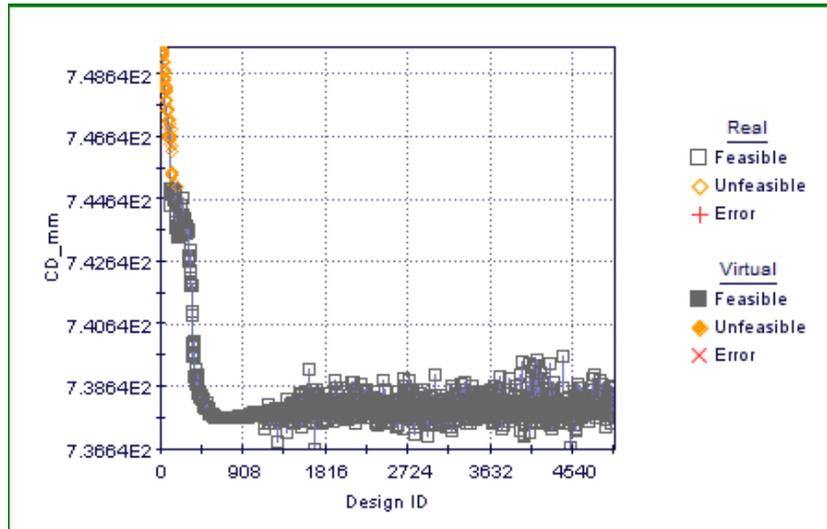


Figure5. Optimization design status based on reduced design domain

图 5. 基于缩减域的优化设计状态

如图 3、图 4 所示，优化设计结果表明，以原始域和缩减域分别作为设计空间的优化设计均得到了收敛的有效优化解，根据两组结果的对比，前者得到收敛的有效解的迭代次数为 294，后者则为 169，其收敛效率提高了近 1 倍，并且根据两个响应变量的约束函数，也可以得出后者得到的解更优。

通过工程实例表明基于决策树的设计空间优化可以有效缩减优化设计域，使出初始设计点最大限度地分布在缩减可行域边缘或者在可行域内，可以避免浪费计算资源，从而有效缓解优化算法寻优复杂度，提高优化算法的寻优效率。

4 结论

本文提出了基于决策树数据挖掘技术优化设计空间的方法，并利用工程实例对该方法的有效性和可行性进行了研究，结果表明所提出的方法实现了对设计空间进行缩减与优化的目的，有效缓解了优化算法寻优的复杂程度，提高优化算法寻优的收敛效率，且能得到更为理想更为优化的有效解。

参考文献 (References)

- [1] Bates RA, Wynn HP, Fraga ES (2007) Feasible region approximation: a comparison of search cone and convex hull methods. *J Eng Optim* 39(5):513-527
- [2] Wang K, Salhi A, Fraga ES (2004) Process design optimization using embedded hybrid visualization and data analysis techniques within a genetic algorithm optimization framework. *Chem Eng Process* 43(5):657-669
- [3] Kuss M, Rasmussen CE (2005) Assessing approximate inference for binary gaussian process classification. *J Mach Learn Res* 6:1679-1704
- [4] Basudhar A, Missoum S, Sanchez AH (2008) Limit state function identification using support vector machines for discontinuous responses and disjoint failure domains. *Probabilist Eng Mech* 23:1-11
- [5] Breiman, L. *Classification and regression tree[M]*. CRC press, 1984.