

基于 MPC 编码方式的软件产品线 配置优化算法研究

郑炜¹, 曹立鑫¹, 李隆俊¹, 李宁²

(1.西北工业大学 软件与微电子学院, 陕西 西安 710072; 2.西北工业大学 计算机学院, 陕西 西安 710072)

摘要:软件产品线配置问题是一个多目标选择难题,借助于遗传算法的全局搜索能力可以得到成本低、耗时少、功能健全的最优方案解。合理的软件产品线特征模型映射编码可以提高求解效率,增加有效解的个数。传统的直接编码是对所有特征进行编码,这使得大型软件产品线配置问题求解效率低下,并且往往得到无效解。现行的强制编码通过隐藏强制节点来缩小特征编码范围,从而达到提高求解效率的目的。然而,很多大型软件产品线配置问题依然未得到有效解决。针对这一问题,提出一种新型编码方式——MPC 编码,在强制编码的基础上,通过节点子父关系进一步缩小编码范围,更加有效提高求解效率,从而获取最优方案解。最后通过传统模型与随机模型进行编码方式验证,将 MPC 编码与直接编码以及强制编码进行对比,证明 MPC 编码在求解软件产品线配置问题中的有效性。

关键词:软件产品线;软件配置;多目标优化;遗传算法;特征模型;编码方式

中图分类号:TP311.5 **文献标志码:**A **文章编号:**1000-2758(2016)01-0176-07

软件产品线是一组具有共同体系构架和可复用组件的软件系统,它们共同构建支持特定领域内产品开发的软件平台。软件产品线集中体现一种大规模、大粒度软件复用实践,是软件工程领域中软件体系结构和软件重用技术发展的结果^[1]。软件产品线特征建模是呈现组件模块之间依赖、互斥等关系的重要手段,自 Kang 等^[2]在 1990 年介绍面向特征域的分析方法提出后,特征建模就被广泛运用于软件产品线中。合理的配置特征模型直接影响和决定了整个产品的质量,然而产品线配置解决方案域并不唯一,因此如何获取最优配置方案解变得尤为重要。

遗传算法是一种求解问题的高度并行性全局搜索算法,它能在搜索过程中自动获取和积累有关搜索空间的知识,并自适应地控制搜索过程以求得最优解,因此它在多目标优化问题的优化求解中有很大的作用和优势^[3]。现阶段研究表明,将软件产品线特征模型映射编码后,通过遗传算法可以有效得到最优配置方案解^[4]。本文将对现有的编码方式

予以改进,通过合并相互依赖性高的模块,进一步合理缩小软件产品线特征模型,从而提高获取最优方案解的效率与个数。经过大量实验证明,这种编码方式是有效的。

1 软件产品线介绍

1.1 特征模型

软件产品线工程是一个系统的发展软件系列产品的方法^[5],其中软件产品被定义为一个特征集合,一个特征就代表着产品的一项功能。特征模型通常被用来表示一个软件产品线中的全部产品^[2],特征模型可以表示为可见的树状结构,树的节点表示特征,树状结构将特征节点相关联。只要遵循特征模型的依赖、互斥等关系进行配置,就可以得到一个有效的软件产品配置。例如,图 1 为 GPS(全球定位系统)的特征模型,其根节点特征为 GPS 产品线本身。根据父节点与子节点之间的关系,其他节点可分为以下几种类型:

1)强制型 定义该节点为其父节点必要节点,标示为●。如图 1 中,寻路系统及人机交互是 GPS 产品线必须实现的功能,屏幕是人机交互必须具备的配置。

2)可选型 定义该节点为其父节点的可选节点,标示为○。如图 1 中,提供路况信息是 GPS 产品线的可选功能。

3)组间或类型 定义为其父节点有且仅有一个子节点,标示为 $g[1,1]$ 。如图 1 中,屏幕必须在触摸屏与液晶显示器中二选一。

4)组间异或类型 定义为其父节点至少有一个子节点,标示为 $g[1,*]$ 。如图 1 中,寻路系统在 3D 地图与自动寻路中至少实现一个功能。

此外,特征之间还存在 2 种约束关系:

1)等价关系 选择特征 A 则并需选择特征 B,表示为单箭头虚线。如图 1 中,路况信息与自动寻路为等价关系。

2)互斥关系 特征 A 与特征 B 不能同时存在,表示为双箭头虚线。如图 1 中,键盘与触摸屏为互斥关系。

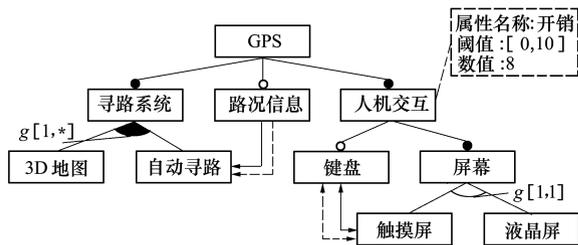


图 1 GPS 特征模型

与此同时,特征模型通过特征属性来扩展特征的额外信息。特征属性的组成通常为:属性名称、阈值与数值。如图 1 中人机交互的一个属性,可知其开销为 8。特征属性不仅用来统计整个产品的细节信息,更是合理配置产品功能的重要依据。

1.2 多目标优化

软件产品线最优化问题主要包含 2 类目标函数:①尽可能少的成本开销;②尽完备的实现功能。此问题也被称为多目标优化问题,多目标优化旨在找出一组同时满足所有优化目标的解集,再由相关决策做出最终选择。传统的多目标优化问题解决方案主要有 2 个:①将多目标转换为单目标,即把各目标加权,然后选择权值最高的解作为最优解;②仅针

对多目标中的某一目标求解,其他目标只需达到一定范围即可。

然而,传统的多目标优化具有很大的主观性、不一致性,并不能有效的得到最优方案解。通过遗传算法可以很好解决这些问题,遗传算法可以并行的处理一组可行的解集,能在一次运算过程中找到满足多个目标的 Pareto 最优解集。

2 新型的编码方式

软件产品线编码方式是将其特征模型依照特征进行编码,使它的排列形式符合遗传算法的基因序列,随后根据特征属性进行选择运算,最终获得软件产品配置最优方案解集。优秀的编码策略能极大的提高获取最优方案解集的效率 and 个数,文章将介绍三种编码方式:直接编码方式,强制编码,以及一种新型的编码方式——MPC 编码。

2.1 直接编码

直接编码即为直接按照层次遍历顺序对特征树进行编码。令特征集为 S ,编码特征集为 S' ,则 $S' = S$ 。例如,图 2 为任意特征模型树,此特征模型的直接编码如图 3 所示。

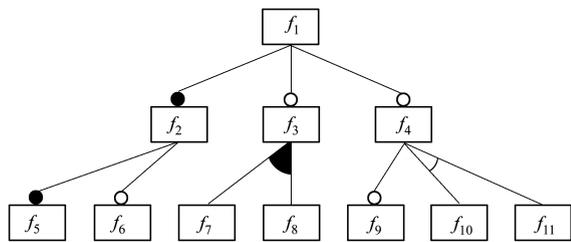


图 2 特征模型树

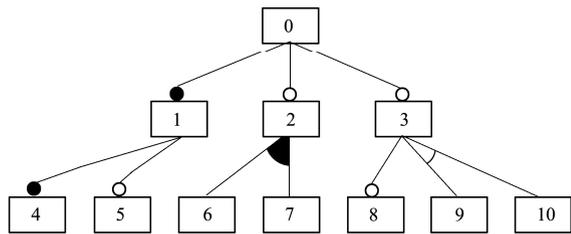


图 3 直接编码

2.2 强制编码

强制编码是将强制型节点和其父节点结合为 1 个特征节点的编码方式。令特征集为 S ,编码特征集为 S' ,强制型节点集为 M ,则 $S' = S - M$ 。例如,在

图 2 中, f_1 存在时 f_2 必定存在, f_2 存在时 f_3 必定存在, 因此 f_1, f_2 与 f_3 是共存的, 于是可将 3 个特征结合为 1 个特征再依照层次遍历进行编码, 如图 4 所示, 从而达到减小模型编码序列的目的, 提高求解配置最优解的效率。

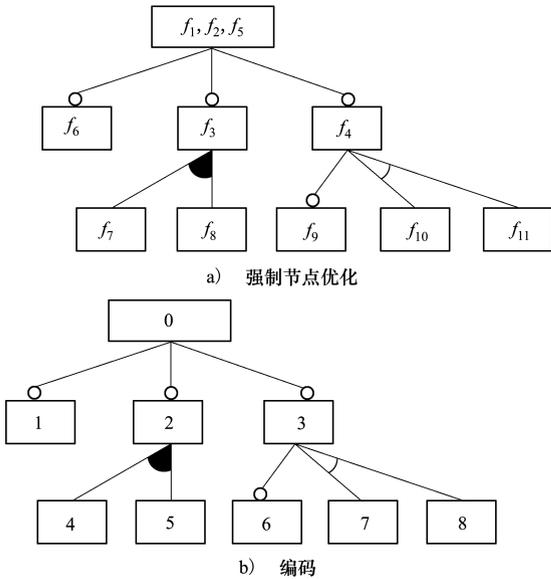


图 4 强制编码

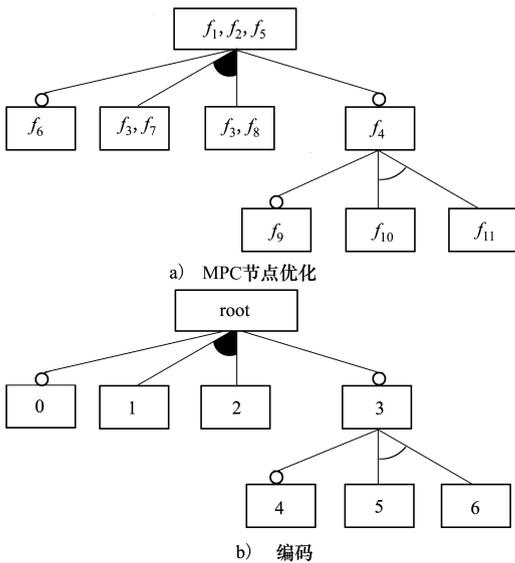


图 5 MPC 编码

2.3 MPC 编码

MPC 编码又称强制子父节点编码, 是基于强制编码改进的编码方式, 通过隐藏子父关系节点达到进一步缩小编码范围的目的。令特征集为 S , 编码特征集为 S' , 强制型节点集为 M , 根节点为 r , 组间型

节点 g 的父节点为 g_f , 该父节点 g_f 的非组间型节点为 f_c , 则:

$$S' = S - M - r - \forall (g_f \cap \neg f_c)$$

例如, 在图 2 中, f_3 有且仅有一组异或子节点 f_7, f_8 , 因此将 f_3 分别与 f_7, f_8 相结合, 然后将其祖父节点变为父节点, 如图 5 所示。MPC 编码进一步减小模型编码序列, 大量实验证明, 该编码有效提高了求解最优配置解的效率与个数。

3 核心算法

文章将采用遗传算法来求解软件产品线最优配置问题, 在算法实现上, 首先将特征按照顺序排列为一个 0、1 序列, 其中序列第 n 位对应特征编码号 $n - 1$, 当序列第 n 位值为 1 时, 表示第 $n - 1$ 个特征被选中, 当序列第 n 位值为 0 时, 表示第 $n - 1$ 个特征未被选中。其次, 若要得到一个合理、可行的产品配置, 需要满足以下 5 个规则:

- 1) 根节点是必要的, 即产品线本身必定存在;
- 2) 任意子节点存在时, 其父节点必定存在;
- 3) 若某节点存在强制型子节点, 则当该节点存在时, 其强制型子节点必定存在;
- 4) 若某节点存在组间子节点时, 则必须符合组间节点定义;
- 5) 必须符合等价关系与互斥关系。

因此, 我们可根据以上 5 个规则生成 1 个规则函数, 通过算法 1 来检验特征序列是否满足配置要求。

算法 1 规则函数

输入: 特征序列

算法步骤:

- Step1 若根节点为 0, 违反规则数+1;
- Step2 若子节点为 0, 父节点为 1, 违反规则数+1;
- Step3 若父节点为 1, 其强制子节点为 0, 违反规则数+1;
- Step4 若父节点为 1, 其组间异或子节点都为 0, 违反规则数+1;
- Step5 若父节点为 1, 其组间或子节点全部为 0 或有 2 个以上为 1, 违反规则数+1;
- Step6 若某节点为 1, 其等价节点为 0, 违反规则数+1;
- Step7 若某节点为 1, 其互斥节点为 1, 违反规

则数+1。

输出:违反规则数。

最终,当得到的违反规则数等于 0 时,表示当前特征序列满足要求,是一个有效的配置方案解。

4 实验配置

4.1 遗传算法

文章将选用 2 种遗传算法来对特征编码方式予以验证: NSGA-II^[6] (精英保留的非劣排序遗传算法) 和 IBEA^[7] (基于指标的遗传算法)。其中, NSGA-II 是 NSGA^[8] (非劣排序遗传算法) 的改进算法,有效地将 NSGA 的计算复杂度从 $O(mN^3)$ 降至 $O(mN^2)$,从而进一步提高算法的计算效率和鲁棒性。IBEA 是一种通用多目标遗传算法,该算法首次将评价指标函数作为适应度评价方法嵌入到遗传算法中,使其可以与任意服从 Pareto 优胜规则的指标相结合,常用的指标有二元 additive epsilon 指标和二元 hypervolume 指标。

4.2 SPL 特征模型

文章选用 2 个传统模型 E-Shopping^[9] 和 Web-Portal^[10],以及 2 个大数据随机模型来验证 3 种编码方式,并使用 Hypervolume 指标(简称 HV)来评估运算结果。E-Shopping 是一个商务网站模型,Web-Portal 是一个门户网站模型,2 个大数据随机模型分

别采用 5 000 特征节点和 10 000 特征节点。其中,每个特征模型的基本数据如表 1 所示。

在实验中,模拟了 5 个目标来定义特征模型,并赋予初值:

1)违反规则数:通过特征模型树所生成的符合 5 个规则的规则函数,计算遗传算法个体解违反规则的数量,取最小值;

2)特征个数:个体解中包含特征的个数,取最大值;

3)特征被用性:个体解中的特征是否曾经被使用过,未曾使用过的特征存在未知缺陷,因此取最大值;

4)已知缺陷个数:曾经被使用过的特征都存在已知的缺陷个数,其值取最小值;

5)开销:每个特征都存在开销,其值取最小值。

5 实验结果及结论分析

在实验中,首先对 4 个特征模型进行编码,编码后的节点个数如表 2 所示。其中,EN 表示编码节点个数,OP 表示优化概率。优化概率的公式为:令特征个数为 f ,编码个数为 e ,则 $OP = (f - e)/f \times 100\%$ 。通过编码结果不难看出,MPC 编码进一步缩小了模型大小,在 10 000 节点随机模型中可优化比高达 40%,使得求解效率得到进一步提高。

表 1 特征模型的基本数据

软件产品线	E-Shopping	Web-Portal	Large Data 5 000	Large Data 10 000
特征个数	290	43	5 000	10 000
强制型节点个数	66	8	1 655	3 209
可选型节点个数	74	17	1 685	3 343
组节点个数	149	17	1 659	1 087

表 2 3 种编码节点统计

软件产品线	E-Shopping		Web-Portal		Large Data 5 000		Large Data 10 000	
	EN	OP/%	EN	OP/%	EN	OP/%	EN	OP/%
直接编码	290	0	43	0	5 000	0	10 000	0
强制编码	215	25.86	35	18.60	3 345	33.1	6 791	32.09
MPC 编码	202	30.34	28	34.88	3 106	37.88	5 922	40.78

表 3 E-Shopping 数据测评结果

编码方式	算法	5 个等价目标的运算结果			限定违反规则数的运算结果		
		HV	VN(/30)	VR/%	HV	VN(/30)	VR/%
直接编码	NSGA-II	0.000 000	0	0.00	0.136 545	13	100
	IBEA	0.000 000	0	0.00	0.169 191	16	100
强制编码	NSGA-II	0.003 343	28	2.07	0.149 158	30	100
	IBEA	0.267 410	30	33.91	0.175 422	30	100
MPC 编码	NSGA-II	0.003 751	30	4.97	0.162 943	30	100
	IBEA	0.256 768	30	84.26	0.190 496	30	100

表 4 Web-Portal 数据测评结果

编码方式	算法	5 个等价目标的运算结果			限定违反规则数的运算结果		
		HV	VN(/30)	VR/%	HV	VN(/30)	VR/%
直接编码	NSGA-II	0.012 498	22	1.52	0.222 625	30	100
	IBEA	0.300 289	30	61.64	0.260 523	30	100
强制编码	NSGA-II	0.059 950	30	1.70	0.275 413	30	100
	IBEA	0.308 381	30	78.99	0.307 052	30	100
MPC 编码	NSGA-II	0.168 637	30	6.02	0.284 126	30	100
	IBEA	0.318 109	30	97.04	0.318 734	30	100

表 5 数据 5 000 节点数据测评结果

编码方式	算法	5 个等价目标的运算结果			限定违反规则数的运算结果		
		HV	VN(/30)	VR/%	HV	VN(/30)	VR/%
直接编码	NSGA-II	0.000 000	0	0	0.000 000	0	0
	IBEA	0.000 000	0	0	0.000 000	0	0
强制编码	NSGA-II	0.000 000	0	0	0.000 000	0	0
	IBEA	0.000 000	0	0	0.000 000	0	0
MPC 编码	NSGA-II	0.000 000	0	0	0.000 652	30	100
	IBEA	0.000 000	0	0	0.000 651	30	100

表 6 大数据 10 000 节点数据测评结果

编码方式	算法	5 个等价目标的运算结果			限定违反规则数的运算结果		
		HV	VN(/30)	VR/%	HV	VN(/30)	VR/%
直接编码	NSGA-II	0.000 000	0	0	0.000 000	0	0
	IBEA	0.000 000	0	0	0.000 000	0	0
强制编码	NSGA-II	0.000 000	0	0	0.000 000	0	0
	IBEA0	0.000 000	0	0	0.000 000	0	0
MPC 编码	NSGA-II	0.000 000	0	0	0.000 000	0	0
	IBEA	0.000 000	0	0	0.000 000	0	0

在求解最优方案解中,2 个遗传算法分别执行 2 遍,1 遍为 5 个等价目标的执行,1 遍为限定违反规定数为 0 的执行。每次执行评估 50 000 次并循环 30 次,即得到 30 组解集。表 3~表 6 为 4 个特征模型的测评结果,其中,VN 表示 30 组解集中含有违反规则数为 0 解的解集个数,VR 表示 30 组解集中所有违反规则数为 0 解所占全部解百分比。

比较实验结果可知,在 3 个编码方式中,MPC 编码的 HV 评估值普遍最优,直接编码的 HV 评估值通常最低,并且,MPC 编码的含有违反规则数为 0 解的解集个数以及违反规则数为 0 解所占比也大部分高于其他 2 个算法。同样地,直接编码普遍偏低。尤其在大型随机模型 LargeData 实验中,直接编码和

强制编码并不能得到有效解,然而 MPC 编码却得到了有效的结果。由此可证实,MPC 编码的确在求解软件配置最优解中更加有效率。

6 结 论

通过大量实验数据结果可验证 MPC 编码方式较优于其他编码方式,然而此方法还有很多不足之处。众所周知,软件产品线特征模型是各式各样的,本方法不一定能囊括所有特征模型,在某些特定模型下并不能绝对提高获取最优方案解的效率。如何使 MPC 编码可以优化更多特征模型是我们进一步研究的问题。

参考文献:

- [1] 李兰涛,王忠民. 基于 UML 的软件产品线建模方法研究[J]. 微计算机信息, 2006, 22(30): 204-206
Li Lantao, Wang Zhongmin. Software Product Lines Modeling Approach with UML[J]. Control & Automation, 2006, 22(30): 204-206 (in Chinese)
- [2] Kang K, Cohen S, Hess J, et al. Feature-Oriented Domain Analysis (FODA) Feasibility Study[R]. Technical Report CMU/SEI-90-TR-21, SEI, 1990
- [3] Goldberg D E. Genetic Algorithms in Search, Optimization and Machine Learning[M]. Boston, Addison-Wesley Professional, 1989
- [4] Abdel Salam Sayyad, Tim Menzies, Hany Ammar. On the Value of User Preferences in Search-Based Software Engineering: A Case Study in Software Product Lines[C]//35th International Conference on Software Engineering, 2013: 492-501
- [5] Clements P, Northrop L. Software Product Lines: Practices and Patterns[M]. Boston, Addison-Wesley Professional, 2001
- [6] Kalyanmoy Deb, AmritPratap, Sameer Agarwal, et al. A Fast and Elitist Multi-Objective Genetic Algorithm: NSGA-II[J]. IEEE Trans on Evolutionary Computation, 2002, 6(2): 182-197
- [7] Zitzler E, Künzli S. Indicator-Based Selection in Multi-Objective Search[C]//Proc of the 8th Int'l Conf on Parallel Problem Solving from Nature, 2004
- [8] Srinivas N, Kalyanmoy Deb. Multi-Objective Optimization Using Nondominated Sorting in Genetic Algorithms[J]. Evolutionary Computation, 1994, 2(3): 221-248
- [9] Sean Quan Lau. Domain Analysis of E-Commerce Systems Using Feature-Based Model Templates[D]. University of Waterloo, 2006
- [10] Marcilio Mendonca, Thiago Tonelli Bartolomei, Donald Cowan. Decision-Making Coordination in Collaborative Product Configuration[C]//Proceedings of the 23rd Annual ACM Symposium on Applied Computing, 2008: 108-113

A New Solution of Configuration Based Genetic Algorithm for Software Product Line

Zheng Wei¹, Cao Lixin¹, Li Longjun¹, Li Ning²

(1. Department of Software Engineering, Northwestern Polytechnical University, Xi'an 710072, China

2. Department of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: A problem, the configuration of software product line, is a puzzle of multi-objective optimization. Optimum solution can be accessed effectively by the capability of searching the optimal solution within defined space form genetic algorithm. Reasonable code of software product line feature model can promote efficiency of global searching and increase the number of efficient solutions. This paper improves current code and obtains a new one - Mandatory Parent Child Encoding. A great number of experimental data indicate that this method is feasible.

Keywords: computational efficiency, cost reduction, genetic algorithms, global positioning system, multi-objective optimization, stochastic models; encoding mode, feature model, MPC (Mandatory Parent Child) encoding, software configuration, software product line