

文章编号 1004-924X(2013)01-0233-06

应用聚类和粗糙集的 FPGA 底层数据挖掘

姜大力^{1,3}, 赵 波², 王 永¹, 李华旺¹, 杨根庆^{1*}

(1. 中国科学院 微系统与信息技术研究所, 上海, 200050;

2. 中国科学院 长春光学精密机械与物理研究所, 吉林 长春 130022;

3. 中国科学院大学, 北京 100039)

摘要:提出了一种从 XDL 文件中提取现场可编程门阵列(FPGA)底层逻辑和布线资源的方法。预处理阶段通过正则表达式将原有底层逻辑文件转换为待分析的有层次属性的关系型数据库。数据挖掘阶段则根据各个层次数据内部的特性不同,采用不同的算法进行聚类来得到初步知识。通过粗糙集分析初步知识间的关系和约简属性,得出初步知识间的联系同时进一步提取出决策规则和产生式规则的知识。最后,通过规则验证器和泛化器对提取出的规则进行验证和泛化。实验结果表明,对于大型的 FPGA 器件,wire 的逻辑最高压缩比可以达到 2.88×10^{-4} 。该方法相对于底层器件有较好的通用性和交互性,适用于对不同器件族 FPGA 底层信息知识提取,对深入研究 FPGA 的拓扑架构,提高对 FPGA 进行动态重配置的可控性和实现更灵活的重配置很有意义。

关键词:现场可编程门阵列;数据挖掘;粗糙集;聚类

中图分类号:TP302.8 **文献标识码:**A **doi:**10.3788/OPE.20132101.0233

FPGA low level data mining based on cluster and rough set

JIANG Da-li^{1,3}, ZHAO Bo², WANG Yong¹, LI Hua-wang¹, YANG Gen-qing^{1*}

(1. *Shanghai Institute of Micro-system and Information Technology, Chinese Academy of Sciences, Shanghai 200050, China;*

2. *Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130022, China;*

3. *University of Chinese Academy of Sciences, Beijing 200050, China)*

** Corresponding author, E-mail: ygq@mail_sim.ac.cn*

Abstract: This paper provides a method to get low level logic and route resources of a Field Programming Gate Array(FPGA) from XDL documents. In the preprocess stage, the lowly logic document is transformed to a relational database with different attributes via a regular expression. In the data mining stage, different algorithms are adopted to cluster the data to get preliminary knowledge based on the different internal features of each level of the database. By taking a rough set to analyze the preliminary knowledge and summarize its attribute, the relation of the preliminary knowledge is obtained and the knowledge for decision rule and production rule is given. Finally, the above rules are validated and generalized through a validator and a generalization device. Experiments show that the highest

收稿日期:2012-09-12;修订日期:2012-10-17.

基金项目:空间科学战略性先导科技专项资金资助项目(No. XD04040202)

compressing rate of wire logic can reach 2.88×10^{-4} for a huge FPGA. This method has good versatility and interchangeability and is suitable for acquiring the low level information and knowledge from FPGAs in different apparatus families. The method has signification in exploring the topology framework of FPGA, improving controllability for dynamic reconfiguration of FPGA, and achieving more flexibility reconfiguration.

Key words: Field Programming Gate Array(FPGA); data mining; rough set; cluster

1 引言

SRAM 型现场可编程门阵列(Field Programmable Gate Array, FPGA)芯片具有发射后可修正与多任务更新等特有优势,可满足空间高可靠性、高性能和多任务的需求,因而在航天领域中越来越受到人们的关注^[1-2]。

SRAM 型 FPGA 的发射后可修正与多任务更新优势主要由可动态重配置的特性来保证。传统的 FPGA 设计流程严重依赖 EDA 软件来完成整个设计流程,这显然与成本高、实时性高的空间设计要求相矛盾。不依赖高层的 EDA 软件,直接对已有的底层逻辑和布线资源进行修改和控制,可以显著地提高配置的可控性,实时性和灵活性^[3-4]。完成这些任务需要有底层资源知识库作为所有功能实现的基础,但这些生产公司或软件研制方的核心技术,只能从工具的介绍中获得底层资源知识库的架构信息,而无法得到底层的知识库表示。

XDL(Xilinx Design Language)是 Xilinx 公司提供的可以访问器件(Device)内所有属性的接口^[5]。XDL 文件有两种语法格式:第一种用来显示器件内的资源信息,通过 XDL 接口可以产生的一个 ASCII 的底层资源描述性文件称为 XDL 信息文件,其中包含了底层硬件的所有信息^[6]。第二种是与器件的设计相对应,每个设计都可以产生一个对应的 XDL 设计文件,同时也可以通过对 XDL 文件的修改和控制原有的设计,实现对已有的底层逻辑和布线资源进行修改和控制^[7]。

通过对 XDL 信息文件的分析可以了解 FPGA 内部的硬件宏逻辑的关系和分布规律,同时指导 XDL 设计文件如何有效修改和控制原有设

计。但 XDL 信息文件只是一个架构性的描述,它以描述性海量数据的方式显示,如 XC2V4000 的 XDL 文件大小超过 1 G,显然不能应用于实际的工程中。

本文将聚类和粗糙集分析等数据挖掘方法应用于 FPGA 底层逻辑资源知识的研究。从海量的 XDL 信息文件中提取 FPGA 底层逻辑和布线资源知识,以提取出的决策规则和产生式规则的知识描述取代原有 XDL 中的海量的数据描述,为实现智能化重配置应用奠定了基础。

2 整体描述

2.1 XDL 文件的语法结构

Xilinx 公司用 Family 表示有相似特性的器件系列,而 Device 表示 Family 内的具体的器件类型。每个确定 Device 都有相应的 XDL 文件与其相对应。XDL 文件的描述中 Device 的描述是以逻辑块的方式来实现,逻辑块命名为 Tile。Tile 是以对称阵列型架构来组织的,整体的 Tile 二维阵列结构被成为 TileMAP。Tile 可以通过网表坐标(row, col)或 Tile 名称来确定,每个 Tile 归属于确定的 Tile 类型(Tiletype)。Tile 内部包含逻辑资源 SLICE 和布线资源 WIRE。

2.2 粗糙集方法

粗糙集作为一种新的处理模糊和不确定知识的数学工具,其主要思想是在保持分类能力不变的情况下,通过知识约简,导出问题的决策和分类规则^[8-10]。XDL 文件的数据挖掘采用了分层次挖掘,采用粗糙集目的在于分析每层知识内部和每层知识之间的粗糙度,并根据不同的粗糙情况采用,不同的规则提取方式。为了分析 XDL 文件,首先将整体 Device 内部

TileMAP 定义为论域

$$TM^j = \{(row, col) \mid row \in \{0, 1, \dots, rows^j - 1\} \wedge col \in \{0, 1, \dots, cols^j - 1\}\}, \quad (1)$$

其中:Tile 以坐标的形式来表示, j 表示 Device 的 id 号, $rows^j$ 和 $cols^j$ 表示 Device 对应的 TileMAP 中的行数 and 列数。通过对各个层次所得到的知识或者聚类集合得到知识或者划分,通过近似精度分析,得到知识间的联系;或者通过集合的上近似分析,收缩集合的定义域,从而提取模式分类规则。

2.3 知识提取方法

XDL 文件的知识提取方法包括 3 个部分,数据预处理,知识提取和规则检验。其中知识提取是整个过程的重点,将在下一节中进行详细介绍。

数据的预处理阶段是通过在 python 中建立与 XDL 语法相对应的正则表达式对 XDL 文件进行语法解析,将原有的层次型数据进行数据结构转换后,通过 python 中的模块将得到的数据存储到对应的待分析对象关系数据库中。

知识提取主要采取的是由顶至底的数据挖掘方式,上一层的数据挖掘作为下一层数据挖掘的先验知识。每一层由于其数据特点和已有的先验知识不同,通过不同方法对属性进行特征提取和聚类,并将结果存入分析数据库。通过粗糙集方法对分析数据库中同一数据层次和层次间的知识进行分析,在进一步属性约简的同时发现属性间的联系形成应用相关的决策规则或产生式规则。

规则验证主要包括规则检验和规则泛化两个部分。规则检验主要通过 python 从待分析数据库中提取假设规则的输入和输出,通过将假设规则的输入传入假设规则,并对假设规则的输出进行比较,统计规则并统计规则的正确率。将正确率大于 95% 以上的规则,不满足规则假设的特例存储在数据库中。规则泛化器则是通过加入特征数据和扩大规则验证范围,直接从原始数据中提取,并不将数据纳入待分析数据库,直接对原始数据进行检验,将规则从局部规则提升到 Device 规则,甚至是 Family 规则。

3 不同层次的数据挖掘概述

3.1 Tile 层数据挖掘

Tiletype 是由语法形成的对 Tile 的基本聚类,在此基础上通过对 Tiletype 进一步挖掘得到 TileMAP 的拓扑规则间的关系。

$$TT^j = TM^j / IND(T) = \{TT_0, TT_1, \dots, TT_i, \dots, TT_{n-1}\}, \quad (2)$$

是以 Tiletype 为等价关系对 TM_{device} 论域的划分,其中 i 为 Tiletype 的 id 号,是 $TM/IND(T)$ 的初等范畴中当等价类 $T = \text{Tiletype}$ 时的元素集合。

规则 1:通过对 Tiletype 在 Device 内分布的图像进行抽样分析,建立假设对于任意的 Tiletype 有

$$ROW_i = \{row \mid row \in (row, col) \wedge (row, col) \in TT_i\}, \quad (3)$$

$$COL_i = \{col \mid col \in (row, col) \wedge (row, col) \in TT_i\}, \quad (4)$$

则有 $ROW_i \times COL_i = TT_i$ 。

通过泛化器,该规则是一个 Family 规则。

规则 2:每个 Tiletype 可以由其他的 Tiletype 来线性表示。由于规则 1 且知识 T 为 Tile 的一个完整划分,Tiletype 间存在着线性相关的关系。因而应该可以通过 Tiletype 中的特征提取,得到产生行列拓扑的特征的 Tiletype。其具体方法如下:

1. 对 Family 中的 Tiletype 在 2 个方向上建立矩阵 R 和 C ,每个 Tiletype 对应在 2 个矩阵中分别对应一个行向量,对于 Tiletype 的 id 为 i 的 Tiletype,其对应的向量分别为

$$\begin{aligned} R_i &= \{cart(ROW_i^0), cart(ROW_i^1), \dots, cart(ROW_i^j), \dots\}, \\ C_i &= \{cart(COL_i^0), cart(COL_i^1), \dots, cart(COL_i^j), \dots\}. \end{aligned} \quad (5)$$

2. 对 R_i 通过 χ^2 相关性分析在 2 个维度 R 和 C 上进行聚类,得到 2 个维度上的相关性等价类 GR 和 GC。每个等价类中的 Tiletype 均保持这线性关系。

3. 对得出的等价类,求秩和 RREF 变换得到最大线性无关组。通过线性无关组的变换可以得到任何一种 Tiletype 的空间大小,行和列的分别展开得到了整个 Tile 的行和列。

3.2 Site 层数据挖掘

Site 层是以 Sitetype 为基本聚类,数据挖掘的重点在于建立 Sitetype 聚类与上一层知识 TT 之间的映射关系,Site 层实例的命名规则,Site 实例内部的 Wire 的命名规则。

首先,建立 Sitetype 和 TM 之间的联系, SI_k 表示 TM 中包含有 Sitetype 的 Site 实例的 Tile 的集合,其中 k 为 Sitetype 的 id。通过粗糙集分析,对于任意的 Sitetype 均有

$$a_T(SI_k) = \frac{|TSI_k|}{|SI_k|} = 1. \quad (6)$$

即 SK_k 是 T 知识可以定义的。由此通过粗糙集得到知识 T 对 SI_k 的表示,并存为关联表。

通过抽样建模得到 Site 实例命名规则,通过 Tile 层知识挖掘中建立的规则 2,将 Site 实例和 Tiletype 的特征值间建立关系,得到 Site 实例的值域空间,由此将 Device 规则提升为 Family 规则。

3.3 Wire 层数据挖掘

Sitetype 和 Tiletype 的映射关系,可以直接用于 Wire 的数据挖掘的分析中,从而只需要对余下的 Wire 进行聚类,聚类的方式是通过 Frag-Shells 方法建立倒排索引,通过粗糙集的方式对得到的聚类结果,与先前得到的知识进行分析,得到知识间的关联,实现了聚类后数据的约简。

4 实验结果

4.1 Tile 层

通过泛化器得到,除特例 XC2V80 中的 EMPTY0X2 外,该规则 1 为 Family 规则,即所有的 Tiletype 均可以映射到一个矩形空间内。通过将原有的 $S(m \times n)$ 的空间改变为 $S(m+n)$ 的空间,然后通过析取的方式得到 Tiletype 和 Tile 坐标间的对应关系。同时,根据实例化数量对进行排序,压缩数据空间的同时提升了平均搜索的期望。

通过 Tiletype 的相关性聚类,66 个 Tiletype 在 RC 2 个方向上各得到 6 个等价类组,对每个方向的组的代表向量求解最大线性无关组,得到 R 方向的 CENTER 和 BRAMSITE_GCLKH 两个 Tiletype,而 C 方向为 CENTER 和 BRAM1,即通

过线性变换可获得每个 Tiletype 映射后的空间及 TileMap 的空间大小。

4.2 Site 层

通过对知识 T 对 SI_k 的关联表的泛化,得出 Device 规则和 Family 规则,并建立相应的 Tiletype 和 Sitetype 间的函数,即通过 Tiletype2Sitetype 函数得到 Tiletype 内含有的 Sitetype,而通过 Sitetype2Tiletype 得到 Sitetype 对应的 Tiletype 组合。

通过值域空间的抽样模式提取,建立了 Site 实例和 Tile 坐标间的函数转换关系。即在确定的 Device 中,通过函数 SiteInst2TileInst 可以通过输入 Site 实例名称,返回对应的 Site 实例所在的 Tile 实例名称。通过 TileInst2SiteInst 函数可以通过输入 Tile 实例名称或者坐标,可以返回 Tile 实例内所包含的 Site 实例的名称。

4.3 Wire 层

对 Wire 进行聚类,得到 53 个类集合。在 TM 论域中,对 53 个集合针对 T 进行粗糙可定义性分析,其中 40 个集合可以用 T 知识来定义 $S_{aD} = \{S | S \in S_T, r_S(T) = 1\}$,而 13 个知识不能通过 T 知识来定义 $S_{aN} = \{S | S \in S_T, r_S(T) = 0\}$ 。对 S_{aN} 进行粗糙集分析,得到 $\forall j, S_{aN_j} \in S_{aN}$,存在 $\bar{T}S_{aN_j} \in S_{aD}$,即任一在 T 内不能定义的 Wire 均为一种在 Tile 类型集合下的特殊情况,且其分组从属与可定义的 S_{aD} 分组集合。由此,每个分组都转换为 Tiletype 的一种映射。

表 1 Wire 压缩情况比较表

Tab.1 Comparison of Wire compression

压缩器件 Virtex II	压缩率	压缩器件 Virtex II Pro	压缩率
2V40	5.0%	2VP ₂	0.9%
2V80	2.5%	2VP ₄	0.42%
⋮	⋮	⋮	⋮
2V6000	2.73×10^{-4}	2VP ₁₀₀	2.88×10^{-4}

表 1 为 Xilinx VirtexII 器件和 VirtexII Pro 器件对于 Wire 知识的压缩比率值。从表中可以看出通过知识挖掘明显压缩了 Wire 的数据

空间。

5 结 论

本文提出了一种从XDL文件中提取FPGA底层逻辑和布线资源的方法。该方法首先通过正则表达式将原有底层逻辑文件转换为待分析的有层次属性的关系型数据库。然后,根据各个层次数据内部的特性不同,采用不同的算法进行聚类从而得到初步知识。通过粗糙集对初步知识间的关系分析和属性约简,得到初步知识间的联系同时进一步提取出的决策规则和产生式规则的知识。最后,通过规则验证器和泛化器对提取出的

规则进行验证和泛化。实验结果表明,通过对XDL信息文件中底层逻辑的数据挖掘,将Tile层中的Tiletype的数据空间从 $S(m \times n)$ 改变为 $S(m+n)$;通过分析Tile层和Site层的关系,同样将原有的描述性规则转换为Tiletype的产生式规则;通过对Wire层的数据挖掘,并建立产生式规则,对于大型的FPGA器件,Wire的逻辑最高压缩比可以达到 2.88×10^{-4} 。通过该方法提取出的决策规则和产生式规则的知识描述取代原有XDL中海量的数据描述,对深入研究FPGA的拓扑架构和提高对FPGA进行动态重配置的可控性,及实现更高灵活性的重配置,有着非常重要的意义。

参考文献:

- [1] MERAZ S, ARVIZU M, CRUZ J M C, *et al.*. GNSS receiver based on a SDR architecture using FPGA devices[C]. *Electronics, Robotics and Automotive Mechanics Conference (CERMA)*, 2011: 383-388.
- [2] IZUMI T, OKADA M, MASAKI T. An FPGA implementation of 200MBPS CI/OFDM modem for satellite communication systems[C]. *World Automation Congress (WAC)*, 2010: 1-6.
- [3] LAVIN C, PADILLA M, LUNDRIGAN P, NELSON B, *et al.*. Rapid prototyping tools for FPGA designs: rapid smith[C]. *International Conference on Field-Programmable Technology (FPT'10)*, December 2010: 353-356.
- [4] STEINER N, WOOD A, SHOJAEI H, *et al.*. Towards an open-source tool flow[J]. *Proceedings of the 19th ACM/SIGDA international symposium on Field Programmable Gate Arrays (FPGA '11)*. ACM, 2011: 41-44.
- [5] Xilinx, Inc. The Xilinx Design Language, Juli 2000, HTML documentation file supplied with ISE Verion 6.3.
- [6] BECKHOFF C, KOCH D, TORRESEN J. The Xilinx Design Language (XDL) - Tutorial and Use Cases[C]. *Proceedings of the 6th international workshop on Reconfigurable Communication-centric Systems-on-Chip (ReCoSoC)*, Montpellier, France, 2011: 1-8.
- [7] GHOSH S, NELSON B. XDL-based module generators for rapid FPGA design implementation[C]. *Field Programmable Logic and Applications (FPL)*, *International Conference*, 2011: 64-69.
- [8] 李杨,徐抒岩,韩诚山,等. 量化故障树分析技术在空间相机软件开发中的应用[J]. *光学精密工程*, 2008, 16(11): 2180-2186.
- LI Y, XU SH Y, HAN CH SH, *et al.*. Application of quantitative fault tree analysis to software development for space camera[J]. *Opt. Precision Eng.*, 2008, 16(11): 2180-2186. (in Chinese)
- [9] 杨兆军,李雪,陈淑芬. 基于粗糙集模糊控制的微孔钻削在线检测[J]. *光学精密工程*, 2008, 16(7): 1230-1235.
- YANG ZH J, LI X, CHEN SH F. On-line monitoring micro-size drilling based on rough set fuzzy control[J]. *Opt. Precision Eng.*, 2008, 16(7): 1230-1235. (in Chinese)
- [10] 张坤华,杨但. 应用聚类和分形实现复杂背景下的扩展目标分割[J]. *光学精密工程*, 2009, 17(7): 1665-1671.
- ZHANG K H. YANG D. Segmentation for extended target in complex back grounds based on clustering and fractal[J]. *Opt. Precision Eng.*, 2009, 17(7): 1665-1671. (in Chinese)

作者简介:



姜大力(1980—),男,吉林长春人,博士研究生,主要从事嵌入式容错设计,FPGA 重配置设计,数据挖掘方面的研究。E-mail: jiangdali@gmail.com



李华旺(1973—),男,江西都昌人,博士,研究员,博士生导师,主要从事现代微小卫星方面的研究。E-mail: sataasic@mail.sim.ac.cn



赵波(1974—),男,吉林长春人,副研究员,主要从事光仪设计,光电跟踪系统方面的研究。E-mail: zb606688@163.com



杨根庆(1952—),男,浙江宁波人,博士,研究员,博士生导师,主要从事现代微小卫星方面的研究。E-mail: ygq@mail.sim.ac.cn



王永(1985—),女,河南周口人,工程师,主要从事卫星可重构计算机软硬件研究。E-mail: smart_witch@sina.com

(本栏目编辑:李自乐)
(版权所有 未经许可 不得转载)