

基于 GainRatio 降维算法的流量聚类研究*

高锐, 刘北水, 李丹, 刘杰, 尤博

(工业和信息化部电子第五研究所, 广东 广州 510610)

摘要: 随着网络数据流量的快速增长, 需要高效的流量分类技术来实现网络管理、流量控制和安全检测。传统基于端口和有效负载的流量分类方法准确率低, 无监督学习方法往往仅采用单一的聚类算法对数据进行聚类分析, 且较少研究对数据本身的处理。为了解决上述问题, 提出了先运用 GainRatio 信息增益率方法对原始数据进行降维处理, 再将降维后的数据进行聚类的方法。实验结果表明: 提出的方法不仅有效地提高了运行效率, 而且随着聚类个数的增加, 也明显地提高了高准确率收敛速度。

关键词: 机器学习; 流量聚类; 网络安全; 维度下降; 信息增益

中图分类号: TP 301.6

文献标志码: A

文章编号: 1672-5468 (2020) S2-0051-05

doi:10.3969/j.issn.1672-5468.2020.S2.017

Traffic Clustering Research Based on GainRatio Dimensionality Reduction Algorithm

GAO Rui, LIU Beishui, LI Dan, LIU Jie, YOU Bo

(CEPREI, Guangzhou 510610, China)

Abstract: With the rapid growth of network data traffic, efficient traffic classification technologies are required to implement network management, flow control and security detection. The traditional port-based and payload-based classification methods have low accuracy, and the unsupervised learning method often adopts only a single clustering algorithm to cluster the data. To solve problems mentioned above, a method of reducing the dimensionality of the original data by using the GainRatio information gain rate method and then clustering the dimensionality-reduced data is proposed. The results show that the proposed method not only effectively improves the operating rate, but also accelerates the convergence rate of high accuracy with the increase of the number of clusters.

Key words: machine learning; traffic clustering; network security; dimensionality reduction; information gain ratio

0 引言

当前, 网络中充斥着大量的流量数据, 新应用层出不穷, 例如: WWW、P2P、多媒体和网络攻击数据等。我们需要高效的网络流量分类技术来实现网络管理、流量控制和安全检测。基于端口或基于有效负载的分类等方法都有自身的缺点, 基于端口的方法在判断应用程序流量方面并不十分准确, 例如: P2P 使用的就是随机端口 (动态端口), 端口号并不固定。基于有效负载的方法在应用方面有很大的局限性, 由于某些应用程序使用加密技术, 所以导致该

方法通常也无法解密数据内容并提取相关的功能字段。即使能提取到相关数据, 也会涉及数据安全问题。另外, 随着新应用的出现, 这种分类方法需要不断地解析应用层协议, 这将消耗大量的计算机资源。

因此, 研究人员开始采用基于流统计特征的有监督学习算法, 即人工将一组包含若干属性特征的数据分类并标记作为训练数据, 通过有监督算法学习训练数据, 然后再将未知流数据输入到已经训练好的分类器, 最终由分类器判断输入的数据类型。上述过程的难点在于事先要花费大量的时间去分类并标记训练数据。在如今的网络大数据环

★基金项目: 2018 年工业转型升级资金项目-信息编码核心算法检测评估能力建设; 广州市科技计划一般项目 (201804010316); 国家重点研发计划项目 (2019YFC0118800); 国家重点研发计划项目 (2018YFC1201104) 资助。

收稿日期: 2020-04-21

作者简介: 高锐 (1986-), 男, 山西朔州人, 工业和信息化部电子第五研究所信息安全中心工程师, 硕士, 主要从事密码与网络安全方面的研究工作。

境下,数据日新月异,单独用人工去逐条分析流数据类型已经显得捉襟见肘。

为了解决上述问题,我们引入聚类。聚类分析是将一组未知数据对象划分为若干个子集的过程。每个子集都是一个集群,这样集群中的对象彼此相似,但与其他集群中的对象不同,由聚类分析产生的聚类集可以被称为聚类。我们可以将未知数据根据聚类算法生成若干分组,然后选择准确率高的聚类方案对数据进行标记。再以此作为训练数据的训练分类器,充分地运用到流量实时分类当中。

有许多研究已经应用聚类算法进行网络流量分类,例如:K-Means、期望最大化(EM)、FarthestFirst、Canopy和鲁棒统计流量分类(RTC)。RTC是将监督学习与无监督学习进行了结合。K-Means、FarthestFirst和Canopy都属于硬聚类算法,每个对象会被准确地分配给一个聚类。相比之下,EM是一种软聚类算法。它只给每个对象到相应类提供一个概率。实际上,它使用最大似然估计法来迭代地寻找最佳概率分布^[1]。

本文主要讨论硬聚类。FarthestFirst、K-Means和Canopy这3种算法在形式上相似,所有的算法都在寻找聚类中心,但每种算法寻找中心的方式有所不同。K-Means和FarthestFirst都需要提前指定聚类中心的数量,但在给定中心方面两者有区别。前者是随机给定中心,然后迭代地重新定义中心,该中心点是质心而不是特定点;后者是将与当前中心设置的距离最远的点视为下一个中心,这是一个实际存在的点。Canopy寻找中心的方法与上述两种方法有很大的不同。首先,随机确定一个点;然后,设置两个阈值 T_1 和 T_2 ($T_1 < T_2$),小于 T_2 的点包含在群集中,大于 T_1 但小于 T_2 的点距离仍然可以是下一个群集中心或成员。两个Canopy群集可以有数据重叠,这是它的一个重要特征。

为了进一步地研究如何利用特征选择来降低维度对各种算法的影响,这里引入一种GainRatio方法。当中采用了经典的迭代二叉树3代(ID3)算法。通过Ranker Search算法评估所有的相关特征,并选择使增益达到最大的特征作为决策树节点,以此类推递归建立决策树分支。

在实验中,我们使用的是Moore等人采集的数据集。为了便于评估聚类效果,我们分别从DataBase、Mail、Server、WWW、P2P和Attack的应用中随机地选取了1 000个对象,每个文件包含用于实验的6 000个对象。同时为了增加实验的可信度,我们随机地抽取了10次总共60 000个对象用于交叉比对。实验结果表明,引入特征选择方案之后,全部聚类算法的表现都得到了很大的提升,就连原本准确率最差的FarthestFirst算法也奇迹般地连续超过Canopy、KMeans算法,这是一个惊人的发现。在后续研究中我们会增加特征选择算法的个数,同时加大每个文件中各个应用的数据规模来充分地讨论不同样本规模、不同特征选择算法对各种聚类算法准确率上的影响。

本文的贡献如下:

- 1) 首次将GainRatio降维算法应用到流量聚类领域;
- 2) 首次将Canopy聚类算法应用到流量聚类领域;
- 3) 使用降维算法处理数据集来提高聚类的准确率,同时提升算法的执行效能。

1 相关工作

Erman等人^[2]第一次将K-means和具有噪声的基于密度的聚类方法(DBSCAN)算法应用于网络流量分类。实验结果表明,与AutoClass相比上述两种算法都具有优异的聚类效果,而且可以更快地收敛。另一方面,尽管DBSCAN的准确度低于K-Means,但它可以产生更好的聚类。

王等人^[3]提出了约束聚类,主要是将一些背景信息作为观察流量统计中的约束条件。研究表明,与以前简单的EM和K-Means算法相比,运用该算法不仅提高了流量聚类的质量,而且加快了聚类分析的收敛速度。

Sharmila等人^[4]发现FarthestFirst算法非常适合大数据集,但不太利于生成统一的聚类器。通过分析不同质量的样本,该算法不受初始值和杂乱数据集的影响。

张等人^[5]结合有监督和无监督学习提出了一种鲁棒网络流量分类(Robust Network Traffic Classification)。它通过从未标记的网络流量中提取零日流量样本并自动优化RTC方案参数来解决零日应用程序问题。

Williams等人^[6]对5种有监督学习算法C4.5、贝叶斯网络(Bayes Net)、朴素贝叶斯离散化(NBD)、朴素贝叶斯核心密度估计(NBK)和朴素贝叶斯树(NB Tree)进行了系统比较。在经过特征选择算法降维后,NBD和NBK算法的准确性变化最大,减少了2%~2.5%。无论是否降维,C4.5正常分类速度最快,适合在计算性能方面的实时分类任务。

Jamuna等人^[7]研究了大量的算法,表明机器学习可以很好地应用于网络流量分类。许多算法显示出了较好的准确率。作者还指出,目前机器学习算法仍然依赖于对抓取流量的离线分析,后面会有越来越多的研究开始解决实时流量分类问题。

2 实验方法

本实验处理过程主要分为两个部分。第一部分为数据降维处理,因为我们随机采取的数据量样本较大且分布不均匀,所以GainRatio信息增益率方法正好能够满足这一要求。

聚类算法主要分为硬聚类和软聚类两种类型。其中,硬聚类表示算法中所有的流数据都将会被准确地分配到一个聚类中。而软聚类则表示将每条流数据按照一定的概率分到各个类中,例如:EM、Fuzzy c-Means。本文主要探讨K-Means、Canopy和FarthestFirst 3种硬聚类算法在数据进行降维处理后聚类准确率的变化。

2.1 Gainratio

要了解GainRatio,首先需要知道什么是信息增益(Information Gain)。Shannon定理指出信息熵表示不确定度,当数据处于均匀分布的时候,信息熵值越高表示数据的不确定度越高。当选择指定的特征对数据集进行相应的分类后,其信息熵会减少,那么减少的信息熵即为信息增益。信息增益可以用来衡量数据中某一特征对分类影响的大小,其值越大说明分类纯度越高,效果越好。这里采用的是由Ross Quinlan发明的迭代二叉树3代(ID3)的算法^[16]。其核心思想是以信息增益来度量选择属性,将属性分裂后

信息增益最大的属性进行分裂。下面给出算法过程^[7]。

假设 H 是一个包含 N 个样本, k 个类的样本集。那么我们可以定义如下信息熵:

$$I(H) = -\sum_{i=1}^k P_i \log_2 P_i \quad (1)$$

式 (1) 中: P_i ——数据集 H 中属于类 i 的概率。

假设 F 是样本数据中的一个属性特征, 它有 m 个不同的取值, 即可以将数据集 H 分为 m 组。那么根据 m 个取值分类的情况再次求数据的信息熵可以有:

$$I(F) = -\sum_{i=1}^m \frac{N_i}{N} \cdot I(H) \quad (2)$$

式 (2) 中: N_i —— k 个类 F 属性的第 i ($i=1, 2, 3 \dots m$) 个取值的样本数量。

那么结合式 (2) 和欧氏距离公式我们可以很容易地得到属性特征 F 的信息增益为:

$$GA(F) = I(H) - I(F) \quad (3)$$

以上信息增益的结果偏向于处理大量数据的情况, 但是, 如果数据分布不均匀则会出现过拟合情况。所以 GainRatio 算法引用了一个分裂信息, 如公式 (4) 所示:

$$\text{SplitInfo}_F(H) = -\sum_{j=1}^m \frac{|H_j|}{|H|} \log_2 \left(\frac{|H_j|}{|H|} \right) \quad (4)$$

那么我们可以将增益比定义如下:

$$\text{GainRatio}(F) = \frac{GA(F)}{\text{SplitInfo}_F(H)} \quad (5)$$

实际上 GainRatio 是一种基于决策树的特征选择算法。在构建决策树过程中需要采用相关的搜索算法来寻找最好的子集。比较常用的有 Greedy search、BestFirst search 和 Ranker 3 种搜索算法。本文实验中采用的是 Ranker 搜索算法, 它不是属性子集的搜索方法, 而是各个属性的排序方案。它通过各自的评估对属性进行排序, 需要结合 GainRatio 一起使用。Ranker 不仅对属性进行排名, 还通过删除排名较低的属性来执行属性选择。我们可以设置一个截止阈值, 低于该阈值的属性将被丢弃, 或者指定要保留的属性个数。另外还可以不考虑属性等级大小去指定必须保留的某些属性。

2.2 K-Means

作为一种知名的聚类算法, K-Means 因其方便、快捷和简单而被广泛地应用于各个领域。Hartigan 等人^[8]提出了关于该算法的实用版本。字母 K 表示我们想要的聚类数。K-Means 的关键是定义质心并计算一个点与另一个点之间的距离, 这不仅适用于一个维度, 还适用于多维度。

欧氏距离^[9]被认为是最常用的聚类方法之一。实际上, 距离与实体之间的相似性密切相关, 即小距离意味着较高的相似性, 大距离表明具有较低的相似性。

质心就是我们上面提到的一个簇中每个点的平均值。首先, 我们将随机选择数据集的 k 个对象作为初始中心。在使用欧氏距离计算中心点的代价后, 剩余点将被分配给最相似的临时组。K-Means 将利用来自每个临时组的对象计算新均值, 而不是诸如初始中心之类的点。因此每个点将根据与新中心的距离重新分配到新的临时组。K-Means 迭代地重复进程, 直到质心不再改变。欧氏距离公式为:

$$E = \sum_{i=1}^k \sum_{x \in G_i} \text{dist}(x, g_i)^2 \quad (6)$$

式 (6) 中: x ——一个聚类中的对象, $x \in G_i$;

G_i ——聚类;

g_i ——聚类 G_i 的质心;

$\text{dist}(x, g_i)^2$ ——每个聚类中的方差;

k ——聚类的数量。

因此我们可以知道 E 是数据集中给出的所有对象的平方差之和。 E 越小, 表示紧凑性越高。

由于初始的随机中心点不同, K 均值算法通常会收敛到局部最优而不是全局最优。为了改善聚类结果, 需要以不同的初始值多次运行 K-Means, 或者将其他算法 (例如 Canopy 算法) 用于微聚类。

2.3 Canopy

Canopy 算法最初在 2000 年被提出^[10]。它主要应用于两个方面: 1) 作为聚类分析中单独的聚类算法, 我们会在后面讨论; 2) 用于 K-Means 算法之前的粗聚类, 它可以有效、科学地帮助 K-Means 获得合理的 k 值 (不合理的值将显著影响 K-Means 的结果), 这极大地加速了 K-Means^[11] 收敛。

Canopy 算法主要通过指定两个阈值 T_1 、 T_2 来将杂乱的数据分组到某些有规则的 Canopy 中, 从而计算从数据到中心的距离。假设有一堆数据散布并放置在集合 S 中, 现在随机选择一个点 P 作为初始中心点, 并设置两个距离阈值 T_1 、 T_2 , 其中 $T_1 < T_2$ 。算法实现如下所述。

输入: 定义 S 为对象集合, C 为存放聚类中心集合, 将 S 第一个元素放入临时聚类中心集合 CN 中; 输出: r 是否加入 C 中。

a) while S 不为空; do 选取 CN 中第一个元素作为中心 c 。

b) while S 没有遍历完; do 计算 S 中剩余点 r 到 c 的距离 $\text{dis}(r, c)$; if ($\text{dis}(r, c) < T_1$), 将 r 从 S 中移除。

c) else if ($\text{dis}(r, c) < T_2$), 将 r 作为下一个中心备选点放入 CN 。

d) 将 r 加入到 C 中。

从上述算法中可以看出, 多个 Canopies 里面的数据可以重叠。因此, Canopy 可能导致较低的精度, 但它非常简单和有效。本文将充分地探讨其在网络流量聚类中的有效性。

2.4 FarthestFirst

FarthestFirst 的核心是 farthest-first traversal 算法, 它是由 Hochbaum 和 Shmoys^[12] 首次提出的, 其目的是为了解决 k 中心问题。在定义代价函数和最大聚类半径的条件下, 最远一次遍历算法可以帮助找到 k 聚类的最佳中心点。

算法的过程类似于 K-Means 算法。它也会找到一个用于聚类的质心, 需要确认 k 值来确定有多少个聚类, 但方法却截然不同。FartherFirst 将选择距离中心点集最远的点作为聚类的中心点, 而不是使用 K-Means 中来自一个迭代组的所有点距离的平均值。确认中心点之后剩余的点将被分配到最近的中心。现在给出该算法的详细描述。

假设 S 是中心的集合, d_i 表示从 P_i ($i \in \{1, 2, 3, \dots, n\}$) 到 S 的距离, 因此 $d_i = \min \{d(P_i, c), c \in S\}$ 。距离 S 最远的点将被选择到下一个中心。该算

法可以通过计算值 d_i 将一个点分配给一个聚类。

3 实验和评估

每个样本文件中的一行数据被称为对象（流）。从计算机网络知识可知，网络交互中的每个数据流将遵循诸如 TCP、UDP 和 FTP 等协议。我们还可以将每个数据流称为包，其中包含基本信息，如源地址、目的地址、源端口号、目的端口号、包大小、包间隔和往返时间等，这些都可以作为数据特征。

本章介绍了聚类的全局准确性，聚类数量从 5, 10, 15...300。操作环境为 Intel Pentium 6200 系列，主频为 2.13 GHz，实用内存为 4 GB。我们的实验主要分为两组，对前 10 个样本通过特征选择算法 GainRatio 进行特征选择。

该过程将选择原始全属性中的一个包含 30 个特征子集。换句话说，每个对象有 30 个维度。第二组使用原始数据集，每个对象包含 248 个属性。为了确保实验的可信度，每种聚类的总体准确度是上述 10 个样本的平均值。在下文中，将首先比较相同维度中的每个算法的准确度和收敛速度。然后，讨论在相同算法下维度变化带来的影响。

3.1 数据集和评估方法

本文中使用的数据集^[13]是由 Moore 等人用高性能检测器^[14]收集的。采集数据集群有连接到某研究机构的大约 1 000 个全双工千兆以太网链路主机，不间断地从 2003 年 8 月 20 日 00:34:21 到 15:22:37 采集超过 377 526 的流量数据。这些流动数据分为 10 个样本，每个样本中有 10 种应用。本次实验将用到其中 6 个应用，每种 1 000 个，每个文件样本 6 000 个对象。

评价聚类结果的关键是计算总体准确率。首先，假设 k 是聚类的数量。对于每个聚类 $j \in (1, k)$ 中，用拥有数量最大的应用类标记聚类 j ，该数量表示被正确聚类的数量，由 TP_j 表示。总体准确率则是所有被正确聚类的总数 TP 除以对象总数 (N)^[15]，即：

$$\text{OverallAccuracy} = \frac{\sum_{j=1}^M TP_j}{N} \quad (7)$$

3.2 3 种算法的准确率比较

全属性情况各算法准确率表现如图 1 所示。从图 1 中可以看出，在没有进行特征选择之前 K-Means 的准确率一直是最高，Canopy 的准确率次之；两种算法随着聚类个数的增多增长率几乎相同，先是急速地增长到 70%，然后平缓地向 95% 收敛。反观 FarthestFirst 算法只在聚类个数为 25 的时候增速较快，超过 25 后准确率增速一直较平缓。而且收敛的结果也不如前两种算法。经过 GainRatio 特征选择算法之后的结果如图 2 所示。在聚类数为 50 之前，3 种算法的准确率优劣如同图 1 中所示一样，但是在特征选择后发生了较大的变化，FarthestFirst 的准确率增长速度奇迹般地在聚类数为 75 的时候超过了 Canopy。结果表明对数据进行相应的维度处理对算法准确率的影响是巨大的。

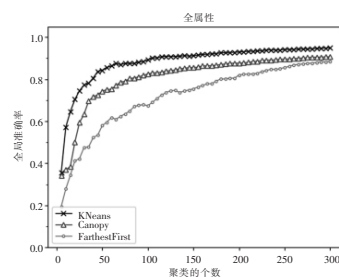


图 1 全属性情况各算法准确率表现

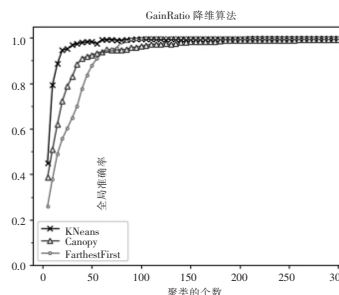


图 2 经过 GainRatio 降维算法后各算法准确率表现

3.3 维度下降给各种算法带来的影响

从图 1-2 中可以知道，随着聚类个数的增加，数据集的维度下降，特别是采用 GainRatio 算法选定了相关特征子集，使得各种聚类算法的准确率都有了很大的提升。3 种算法最后都能很好地收敛到 96% 左右。其中，K-Means 和 Canopy 的准确率变化幅度基本保持一致，而 FarthestFirst 在聚类个数只有 20 个以内的时候准确率增幅并没有明显的差距，在聚类数为 20~80 的区间上面增幅已经接近 100%。

3.4 运行性能

经过 GainRatio 处理后和全特征的数据在 3 种算法中的运行时间比较如表 1 所示。从表 1 中的数据可以看出，所有的聚类算法不管是最小聚类时间还是最大聚类时间，降维后的运行时间都有不同程度的减少。而降维算法的执行时间一般都在 10 ms 左右，对于整体的运行效能影响很小。

表 1 各种聚类算法在降维前、降维后运行时间比较

聚类算法	t/s	
	GainRatio	全属性
K-Means	min: 0.11	max: 1.53
	min: 1.01	max: 11.04
Canopy	min: 0.11	max: 0.13
	min: 0.43	max: 0.76
FarthestFirst	min: 0.08	max: 0.33
	min: 0.11	max: 2.42

4 结束语

首次被运用到网络流量聚类领域，GainRatio 算法表现不俗，降维前后的效果差距比非常明显。实验中 3 种聚类算法的准确率都得到了较大的提升，其中 K-Means 的效果最好，聚类个数只有 25 的时候聚类准确率已经在 95% 以

上,相较于降维前,降维后 Canopy 和 FarthestFirst 算法的高准确率收敛速度非常快。

作为 K-Means 的粗聚类算法,Canopy 也是首次被运用到流量聚类领域,实验结果表明在全特征数据集下准确率超过 FarthestFirst 算法,降维后仍然表现良好,且最后都能收敛到很高的准确率。

同时在算法运行时间上,降维后比降维前平均要快 2~10 倍,每个数据集降维算法的时间基本上可以忽略不计。故可以充分运用到后面的实时流量分类当中。

在未来的工作中,我们将添加更多的聚类算法和特征选择算法。同时,除了数据降维之外,数据体量大小对聚类带来的影响也将是后续研究的重点。

参考文献:

- [1] MCGREGOR A, HALL M, LORIER P, et al. Flow clustering using machine learning techniques [C]. Berlin: Springer-Verlag Press, 2004: 205-214.
- [2] ERMAN J, ARLITT M, MAHANTI A. Traffic classification using clustering algorithms [C]. New York: ACM Press, 2006: 281-286.
- [3] WANG Y, XIANG Y, ZHANG J, et al. Internet traffic classification using constrained clustering [J]. IEEE transactions on parallel and distributed systems, 2013, 25 (11): 2932-2943.
- [4] SHARMILA, KUMAR M. An optimized farthest first clustering algorithm [C]. NJ: IEEE Press, 2013: 1-5.
- [5] ZHANG J, CHEN X, XIANG Y, et al. Robust network traffic classification [J]. IEEE/ACM transactions on networking, 2014, 23 (4): 1257-1270.
- [6] WILLIAMS N, ZANDER S, ARMITAGE G. A preliminary performance comparison of five machine learning algorithms for practical IP traffic flow classification [J]. ACM SIGCOMM Computer Communication Review, 2006, 36 (5): 5-16.
- [7] JAMUNA A, EDWARDS V. Survey of traffic classification using machine learning [J]. International Journal of Advanced Research in Computer Science, 2013, 4 (2): 65.
- [8] HARTIGAN J A, WONG M A. Algorithm AS 136: A k-means clustering algorithm [J]. Journal of the Royal Statistical Society. Series C (Applied Statistics), 1979, 28 (1): 100-108.
- [9] HOLANDA Filho R, MAIA J E B. Network traffic prediction using PCA and K-means [C]. NJ: IEEE Press, 2010: 938-941.
- [10] MCCALLUM A, NIGAM K, UNGAR L H. Efficient clustering of high-dimensional data sets with application to reference matching [C]. New York: ACM Press, 2000: 169-178.
- [11] KUMAR A, INGLE Y S, PANDE A, et al. Canopy clustering: a review on pre-clustering approach to K-means clustering [J]. Int. J. Innov. Adv. Comput. Sci. (IJACS), 2014, 3 (5): 22-29.
- [12] HOCHBAUM D S, SHMOYS D B. A best possible heuristic for the k-center problem [J]. Mathematics of Operations Research, 1985, 10 (2): 180-184.
- [13] MOORE A, ZUEV D, CROGAN M, et al. Discriminators for use in flow based classification [M]. London: Queen Mary and Westfield College Press, 2005: 1-14.
- [14] MOORE A, HALL J, KREIBICH C, et al. Architecture of a network monitor [C]. Berlin: Springer-Verlag Press, 2003.
- [15] WANG Y, XIANG Y, ZHANG J, et al. A novel semi-supervised approach for network traffic clustering [C]. Milan: IEEE Press, 2011: 169-175.
- [16] 李敏, 卡米力·木依丁. 特征选择方法与算法的研究 [J]. 计算机技术与发展, 2013 (12): 16-21.
- [17] KAREGOWDA A G, MANJUNATH A, JAYARAM M. Comparative study of attribute selection using gain ratio and correlation based feature selection [J]. International Journal of Information Technology and Knowledge Management, 2010, 2 (2): 271-277.

基于 EDA 工具的自动化建模平台建设

彭湘涛, 钟日升, 梁仕章, 丘桂全, 李柳琼

(工业和信息化部电子第五研究所, 广东 广州 510610)

摘要: 随着科学技术的发展, EDA 工具在电子电气设计方面越来越不可或缺。EDA 模型作为设计的基础单元, 是非常重要的设计数据, 在当前电子产品需求不断扩大、更新频率持续加快的形势下, 自动化建模已成为提高电子产品的设计效率、最大化降低研制成本的极其重要且有效的实现途径。通过对自动化建模流程进行梳理, 分析出基于 EDA 工具的自动化建模平台的实现方式, 可为广大设计师开展产品电路设计提供足量、准确的 EDA 模型库。

关键词: EDA 工具; EDA 模型; 自动化建模

中图分类号: TP 311.521

文献标志码: A

文章编号: 1672-5468 (2020) S2-0055-04

doi:10.3969/j.issn.1672-5468.2020.S2.018