Contemporary Engineering Sciences, Vol. 7, 2014, no. 8, 363 - 368
HIKARI Ltd, www.m-hikari.com
http://dx.doi.org/10.12988/ces.2014.4220

A Study on Various Data Mining Algorithms Pertaining to VLSI Cell Partitioning

R. Sujitha

School of Computing
SASTRA University, Thanjavur-613401, India

R. Manikandan

School of Computing
SASTRA University, Thanjavur-613401, India

Abstract

VLSI Cell Partitioning plays a substantial part in VLSI physical design. The task of designing integrated circuits is multifaceted, as modern circuits have a very huge number of modules. It is very much essential to split the circuit into a smaller and meeker logic blocks. In order to build a complex digital integrated circuit, the multi million transistors must be sub-divided into convenientnumber of pieces.This paper provides a heuristic approach by utilizing data mining algorithms to solve VLSI partitioning problem. A detailed study on several data mining algorithms like K-nearest neighbour, Support Vector Machine (SVM), Fuzzy c-means and K-means algorithms and their implementation in VLSI Cell Partitioning is provided.

Keywords: Data mining, Partitioning, K-nearest neighbor, Support Vector Machine, Fuzzy c-means, K-means algorithm

1. INTRODUCTION

The most promising outcome of VLSI technology is to develop high-speed, low-power computing structures dedicated for various applications. The design must satisfy numerous constraints on area, power, pins, throughput and noise. One of the major defies in designing a microelectronic VLSI chip is to successfully partition the design into its components[1]. VLSI physical design comprises of three main prospects: Partitioning, Placement, Routing. Division of a complex system design into sub-components,for the purpose of reducing the complexity, is termed as partitioning. Placement involves the process of assigning locations to the components within the die area. The process of interconnecting the placed components, subject to the constraint of minimum wiring, is known as Routing.

Partitioning is a domineering area of VLSI Physical design. The main objective of partitioning is to divide the complex circuit into sub-blocks such that designing these sub-blocks individually and then assembling it separately, reduces the complexity level of the entire design to a much greater extent. Data mining algorithms provide a means to analyse the data and extract information from them. Imbibing data mining approaches into VLSI partitioning pave a new way to research domain.

Data mining algorithms basically are of two types: Classification algorithms and Clustering algorithms. Both these approaches find their utility in recognising hidden patterns in data.Classification involves in forecasting a firm outcome based on a given input. In order to predict the outcome, the algorithm processes a training set that contains a set of attributes and theircorresponding outcome, usually called goal or prediction attribute. The algorithm tries to determinerelations between the attributes that makes it possible to forecast the outcome.

Cluster analysis is demarcated as a process of producing groups of objects, or clusters,in a way that alikeobjects are present in one cluster and objects in unlike clustersare divergent [2].The main rewards of cluster techniques is that it can be applied to a specified objective standard consistently to form groups. The speed and consistency of the clustering procedure in organizing the data together form an irresistible reason to employ clustering in partitioning.

Section II describes the four data mining algorithms in detail. Section III defines the implementation of the data mining algorithms in VLSI cell partitioning. The results are discussed in the section IV.

2. ALGORITHM DISCUSSION

This section briefs about the Classification algorithms: K Nearest Neighbour and Support Vector Machine algorithm and Clustering algorithms: Fuzzy C means and K means algorithm. Though Classification and Clustering algorithms are considered to be ambiguous, their differences are very inimitable. There are predefined set of classes in the classification process and decision has to be made on assigning the already existing class to the new data set. The underlying process in clustering technology is to determine some relation between the data sets and group them based on their relation.

2.1 CLASSIFICATION ALGORITHMS

The main idea behind classification algorithm is selection of a hypothetical set from the existing sets that best fits the given input data set. This section involves detailed description of the selected data mining classification algorithms: K Nearest Neighbour Algorithm and Support Vector Machine Algorithm.

2.1.1 K NEAREST NEIGHBOUR ALGORITHM

K Nearest Neighbour (KNN) algorithm is a lazy learning algorithm.IT is also non-parametric in nature. Assumptions are not done on the underlying data as the practical data does not follow any hypothetical assumptions. There is no overt generalization stage or it is very negligible. On the basis of following presumptions, KNN algorithm is performed. The algorithm presumes that records are in metric space. The data can be scalar type or multidimensional type. Each of the data points has a notion of distance as they are in feature space. The training data set contains a vector set and a class taginterrelatedto the vector. But KNN works well with random number of classes. A single number "k" which denotes the number of neighbours that are defined based on the distances (The distance is calculated in one of the following techniques: Euclidean Distance, Minkowski Distance, Mahalanobis Distance.)[4]. If k=1, then the algorithm is simply called the nearest neighbour algorithm.

2.1.2 SUPPORT VECTOR MACHINE ALGORITHM

A Support Vector Machine algorithm(SVM) constructs a model that assigns given data set into one category or the other, to which the training data sets belong. Thus we can call this SVM algorithm as a non-probabilistic bilinear classifier. Data is represented in spatial domain, so that training data set of different categories are separated by a distinct distance. Therefore the risk of mismatching the given data set is highly reduced. However the SVM algorithm takes a longer time and CPU memory than KNN algorithm. The SVM algorithm suffers a serious setback as it is directly applicable only in two classes and there is an uncaliberated class relationship probability.

2.2 CLUSTERING ALGORITHMS

The idea behind clustering algorithms is to group objects into subsets or clusters based on their relationship with each other [6]. Data in the same cluster will have related properties whereas those in different clusters will have unrelated properties[2]. The algorithms discussed under this section are Fuzzy C means algorithm and K means algorithm.

2.2.1 FUZZY C-MEANS Algorithm

In fuzzy clustering, each data point in the data set is related with every other cluster using a relationship function. The c-means algorithm allows overlapping of dataset, i.e. the given data set can belong to two or more clusters at the same time. Let Y= $\{y1, y2, \ldots, y_M\}$ be a set of numerical data in $R^{N.}$ Let jbe an integer, 1 < j< M. Given X, we say that jfuzzy subsets $\{a_k\text{:}Y \rightarrow [0, 1]\}$ are a j-partition of Yif the following conditions are satisfied[ ]:

$$0 \le a_{kl} \le 1 \text{ for all } k,l. \qquad \ldots 1.A$$

$$\sum_{k=1}^{j} a_{kl} = 1 \text{ for all } l. \qquad \ldots 1.B$$

$$0 < \sum_{l=1}^{M} u_{kl} < n \text{ for all } k \qquad ...1.C$$

where $a_{kl} = a_k(y_l)$, $1 \leq k \leq j$ and $1 \leq l \leq M$. Let the jMvalues a_{kl}filling the above conditions be arrayed as a $j \times M$matrix $A = [a_{kl}]$.Then the set of all such matrices are the non-degenerate fuzzy j-partitions of Y:

$$M_{fcM} = \{A \in R^N : a_{kl} \text{satisfies conditions (1.A, 1.B and 1.C)} \forall k \text{ and } l\}.$$

If all the a_{kl}'s are either 0 or 1, we have the subset of hard j-partitions of Y:

$$M_{cM} = \{A \in M_{FcM} : a_{kl} = 0 \text{ or } 1 \ \forall k \text{and } l\}.$$

This Fuzzy j-partitions is termed as Fuzzy C-partitions.Fuzzy clustering is used in mining complex and multi-dimensional datasets.Fuzzy-C-means (FCM) algorithm iswidespread, in which a portion of data has restrictedrelation with each of the pre-defined cluster centres [5].

2.2.2 K MEANS ALGORITHM

The K means algorithm is designed to cluster data on the basis of cluster centre called mean. The sum of clusters *k* is supposed to be static. The algorithm begins for initial k number of clusters, by assigning the residual data to the adjoining clusters and repetitivelyaltering the relationship of the clusters between each other until the membership becomes stable. It has two stages: Initialization and Iteration stages.During the former phase, the data is allotted into*k* clusters. In the latter phase, the distance between the data setand each cluster is calculated and the data set is assigned to the nearest cluster. However, K means algorithm does not employ overlapping of the data sets.

3. IMPLEMENTATION

The Classification and Clustering algorithms are implemented for the below circuit. The gate at every nodule is represented by the gate matrix, GM, and the assembly of gates is represented by an input matrix, IM. The gate matrix is designed one column at a time, with each consequent columncontaining the gates. The input of the gates may be from the gatesof previous column or wires; therefore thetotal number of columns in GM, C_{GM}, is equal to the maximumnumber of gates on any path from circuit input to output [3].

Table 1. Gate matrix representation

NUMBER	TYPE	NUMBER	TYPE
0	NO GATE	5	WIRE
1	AND	6	NAND
2	OR	7	NOR
3	XOR	8	XNOR
4	INV	9	BUF

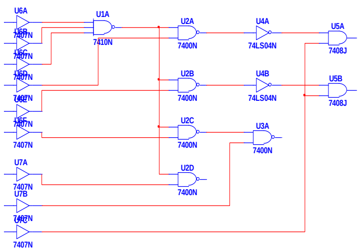
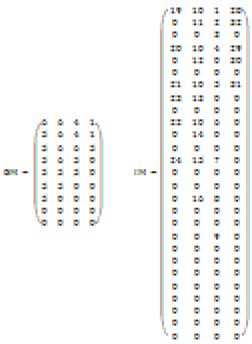


Fig 1. Example circuit

The number of rows in GM, R_{GM}, denotes the cut size of the circuit. The gate in each node is represented by a number in the gate matrix.The corresponding numbers for each gate are itemized in Table 1.The relation between gates in the gate matrix is represented in input matrix, IM, in such a way that the number of columns in input matrix and gate matrix are equal[3].Subsequently, the number of rows ininput matrix, R_{IM}, is the product of $R_{GM}$and the maximum fanin of any gate in the circuit.Therefore, the promising values forcolumn X are $\{1+R_{GM} \times [X-1], R_{GM}\times X\}$ [3]. The four algorithms discussed are applied to the example circuit. The matrix is given as input to the classification and clustering algorithms.



4. RESULT AND DISCUSSION

The Classification algorithms, K nearest neighbour and support vector machine algorithms yield the results as shown in figure 2 and 3.The K Nearest neighbour algorithm gives better classification result than the SVM algorithm as the VLSI cell partitioning employs multiple classes. The SVM algorithm fails to produce efficient result while operating for multiple classes. The Clustering algorithms, Fuzzy C-means and K means Algorithm yield the results shown in figure 4 and 5 when implemented on the circuit. The object function is greatly minimized in the Fuzzy C-means algorithm with less number of iterations when compared to the K means algorithm. Moreover, permitting overlapping of the data sets makes the fuzzy C-means to be more efficient than K means algorithm.

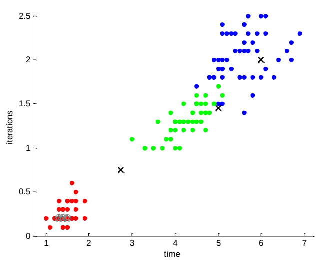


Fig 2.KNN Algorithm

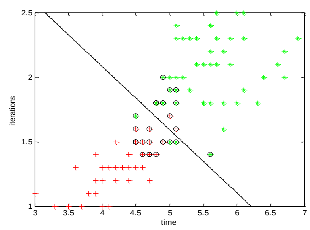


Fig 3. SVM Algorithm

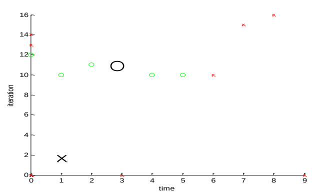


Fig 4. FCM Algorithm

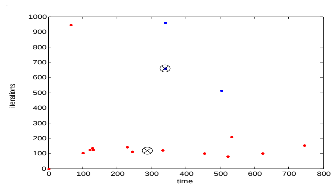


Fig 5.K Means Algorithm

5. CONCLUSION

In this paper, the VLSI cell partitioning which plays a significant role in VLSI physical design is approached with the concept of data mining. The two main data mining approaches, Classification and Clustering are analysed. The Classification algorithms, K Nearest Neighbour and Support Vector Machine algorithms are implemented with the given circuit. It is identified that K Nearest Neighbour gives the optimum solution with regards to the VLSI Cell Partitioning. The Clustering algorithms, Fuzzy C-means and K means algorithms are also implemented on the circuit. It is recognized that Fuzzy C-means algorithm provides optimum solution to VLSI cell partitioning.

REFERENCES

[1] Brian Von Herzen,VLSI Partitioning of a 2-Gs/s Digital Spectrometer,IEEE journal of solid-state circuits, VOL. 26, NO. 5, MAY 1991

[2] Gan, Guojun, Chaoqun Ma, and Jianhong Wu, Data Clustering: Theory, Algorithms, and Applications, ASA-SIAM Series on Statistics and Applied Probability, SIAM, Philadelphia, ASA, Alexandria, VA, 2007.

[3] Ganesh K. Venayagamoorthy et al., Particle swarm-based optimal partitioning algorithm for combinational CMOS circuits,Engineering Applications of Artificial Intelligence (2006), doi:10.1016/j.engappai.2006.06.011.

[4] InduSaini, Dilbag Singh, ArunKhosla.,Detection of QRS-complex using K-nearestneighbouralgorithm., Int. J. Medical Engineering and Informatics, 2013, 5(1),81-101.

[5] SubhagataChattopadhyay, Dilip Kumar Pratihar, Sanjib Chandra De Sarkar, A Comparative study of fuzzy c-means algorithm and entropy-based fuzzy clustering algorithms,Computing and Informatics, 2011,30, 701–720.

[6] Sumithradevi. K.A, Vijayalakshmi.M.N, Optimization of Circuit Partitioning in VLSI through Classification Algorithms, IEEE 2010 4(10).

Received: February 20, 2014