

FPGA Implementation of a Reduct Generation Algorithm based on Rough Set Theory

Kanchan Tiwari¹, Ashwin Kothari² & Riddhi Shah³

¹MESCOE, Pune, ²VNIT Nagpur, ³E&TC

E-mail : kanchan.tiwari@mescoepune.org¹, agkothari72@rediffmail.com², riddhi_shah560@yahoo.com³

Abstract - Rough Set theory, proposed by Pawlak in 1982, is a mathematical tool for classification and analysis of imprecise, uncertain, incomplete or vague information and knowledge. Using various concepts in RST, classification of objects in various classes is done by removing irrelevant and redundant data by using attribute reduction algorithm. Attribute reduction using Rough set theory is one of the key problems. Also to find minimal attributes is a NP-hard problem. Attribute reduction is a preprocessing step in data mining, pattern recognition, machine learning, etc. In this paper, we present a literature survey of various existing algorithms and comparing them on its time and space complexity. In this paper, a hardware implementation of attribute reduction algorithm for classification purpose is proposed. Functional verification of model is done and simulation results are shown.

Keywords - rough set, lower approximation, upper approximation, decision class, time complexity, space complexity, attributes

I. INTRODUCTION

Rough set theory (RST) is an effective tool for mining deterministic rules from a database. It is considered one of the first non-statistical approaches in data analysis. With various concepts of RST like lower and upper approximation, information system etc. objects can be placed in different decision classes as per their features. Reducts can be defined as the information obtained by omitting or neglecting the unwanted data from the system.

Since a decade or more, attribute reduction using RST is a recent and hot area of research for the researchers, since finding reducts is a NP-hard problem. Thus, the main objective of attribute reduction algorithm is to get reducts from the information system by eliminating redundant information and at the same time have minimum time and space complexity. The reducts

obtained should be accurate. The need of finding reducts is to remove redundancy and incompatible attributes to obtain key information and thus make decision rules. If these reducts are not removed, then not only the time and space complexity of rule discovery increases but also the quality of the rules discovered degrades. The cost of finding the reducts depends on the size of object and attributes set. Many attribute reduction algorithms have been proposed till now like heuristic algorithm, algorithm based on discernibility matrix, unsupervised and supervised quick reduct algorithm, algorithm on 0-1 integer programming, etc. Some algorithms proposed are complete while some are not. The time and space complexity of all are different.

Implementation of the reduct generation algorithm is usually simulated using software like RSES, ROSE, ROSETTA etc., but the processing speed using software is not fast enough to meet the demand of real time. The research on its hardware implementation is still an exploring area. It started with an idea put forth by Pawlak himself. He stated that RSP can be used as an additional fast classification unit in ordinary computers or as an autonomous learning machine. Lewis in 1999 identified that the RST algorithms have a natural high parallelism and high possible speed-ups. Using a fast prototyping tool, the DEC-PERLE-1 board based on an array of Xilinx FPGAs, he developed a virtual SIMD processor that accelerates the learning (design) of optimized multi-valued logic nets using the concept of cellular networks. He implemented basic rough set operation of basic category, upper approximation, and lower approximation, indispensable and external comparison. Kanasugi et al initially presented the design and architecture of rough set processor and then implemented his idea of computing reduct and generating rules using Skowron's discernibility matrix concept. They developed a co-processor which will be

interacting with memory for data retrieval and storage purpose. Their co-processor is capable of dealing objects of size 1000,000 and 2032 attributes.

FPGA because of its feature like lower cost, higher density, shorter design cycle, makes hardware design process more flexible and programmable. In this paper, VHDL is used for design and implementation of attribute reduction algorithm. Such chip can be effectively deployed in Robot working at an automated plant assembly. Robot will have sensors mounted on body for capturing features of objects. These features will be processed and a set of minimal feature obtained will be used for making meaningful rules.

In this paper, a literature survey of different proposed algorithm is shown. The algorithms are compared on their space and time complexity. The paper is organized as follows:

1. RST CONCEPTS
2. VARIOUS ALGORITHMS ON ATTRIBUTE REDUCTION
3. SUMMARY
4. CONCLUSION AND FUTURE WORK

II. ROUGH SET THEORY CONCEPTS

Rough set theory works on principle of decision making without requiring human intervention. RST can discover profitable knowledge from incomplete database. In rough set theory, knowledge is look as the classification ability of objects. Some basic definition of rough set theory is given below:-

Definition1.

Let U be a finite set and R be an equivalence relation on U . This relation R will generate a partition $U/R = \{Y_1, Y_2, \dots, Y_m\}$ on U , where $Y_1, Y_2, \dots, Y_m$ are the equivalence classes generated by R [4]. These equivalence classes are also called the elementary sets of R [16]. For any $X \subseteq U$, we can describe X by the elementary sets of R and the following two sets

$R_*(X) = \cup \{Y_i \in U/R: \{Y_i\} \subseteq X\}$, $R^*(X) = \cup \{Y_i \in U/R: \{Y_i\} \cap X \neq \emptyset\}$, which are called the lower and the upper approximation of X , respectively.

The set $BN_B(X) = R^*(X) - R_*(X)$, will be referred to the B-boundary region of X [17].

If the boundary region of X is the empty set, i.e., $BN_B(X) = \emptyset$, then the set X is crisp with respect to $BN_B(X)$; in the opposite case, i.e., if $BN_B(X) \neq \emptyset$, the set X is rough with respect to $BN_B(X)$.

Definition 2.

Information System and Decision Tables

Information is often available in a form of data tables, known as information systems, attribute-value tables or information tables. An information system (IS) is defined as a family of sets $S = \langle U, A, V, f \rangle$, where U is a non-empty universe of objects, A is a finite non-empty set of attributes, V is the value set of A and $f: U \times A \rightarrow V$ is information function. Columns of an information table are labelled by attributes, rows - by objects and entries of the table are attribute values [15]. Objects having the same attribute values are indiscernible with respect to these attributes and belong to the same block of the partition (classification) determined by the set of attributes.

Information systems with distinguished decision and condition attributes are called decision tables. A decision table is denoted by $DT = \langle U, CUD, V, f \rangle$, where U, V and f are the same as defined above; $A = CUD$, C is a set of condition attributes, D is a set of decision attributes, and $C \cap D = \emptyset$. Each row of a decision table determines a decision rule, which specifies decisions (actions) that should be taken when conditions pointed out by condition attributes are satisfied [16].

Definition 3.

Reduct and Core

Every concept in the knowledge base [7] can be only expressed in terms of basic categories. On the other hand, every basic category is built up of some elementary categories. The concepts of core and reduct [9, 11] are two fundamental concepts of the rough sets theory in the case of attributes and knowledge. The reduct is the essential part of which makes all objects discernible by the original information system. The core is the common part of all reducts. The set of all indispensable attributes is called core [15]. Logical rules derived from experimental data may be used to support new reduction. We often face a question whether we can remove some data from a data table preserving its basic properties, that is, whether a table contains some superfluous data.

Definition 4.

Let $F = \{X_1, X_2, \dots, X_n\}$, be a family of sets such that the set $X_i \subseteq U$, $i = 1, 2, 3, \dots, n$. We say that X_i is dispensable in F , if $\cap (F - X_i) \subseteq [X_i]_d$; otherwise the set X_i is indispensable in F . The family F is independent if all of its components are indispensable in F ; otherwise F is dependent. The family $H \subseteq F$ is a reduct of F , if H is

independent and $\cap H \subseteq [X_i]_d$. The family of all indispensable sets in F will be called the core of F .

Definition 5.

$CORE(F) = \cap RED(F)$, where $RED(F)$ is the family of all reduct of F .

Example illustrating concepts of RST:

TABLE I

U	a	b	c	d	e(decision)
1	120	75	Yes	Yes	Accept
2	90	86	No	Yes	Reject
3	100	91	No	No	Reject
4	110	79	Yes	No	Accept
5	115	76	Yes	Yes	Accept
6	109	89	No	No	Accept

Table 1: Non binary information table

TABLE II

U	a	b	c	d	e(decision)
1	1	0	1	1	1
2	0	1	0	1	0
3	0	1	0	0	0
4	1	0	1	0	1
5	1	0	1	1	1
6	0	1	0	0	1

Table 2: Binary conversion of table 1

- Equivalence class (e):
 [1]: {1, 5}
 [2]: {2}
 [3]: {3, 6}
 [4]: {4}
- Class (d):
 X1: {1, 4, 5}
 X2: {2, 3}
- Lower Approximations:
 $\underline{AX1} = \{1, 5\} \cup \{4\} = \{1, 4, 5\}$
 $\underline{AX2} = \{2\} = \{2\}$
- Lower Approximations:
 $\bar{AX1} = \underline{AX1} = \{1, 4, 5\}$
 $\bar{AX2} = \underline{AX2} \cup \{3\} = \{2, 3, 6\}$

- Boundary Region

$$(\bar{AX1} - \underline{AX1}) = 0$$

$$(\bar{AX2} - \underline{AX2}) = \{3, 6\}$$

BLOCK DIAGRAM OF FINDING ATTRIBUTES FROM AN INFORMATION TABLE AND THEN GENERATION OF RULES FROM IT

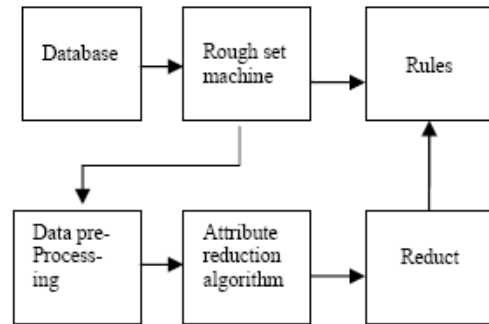


Fig. 1: Block diagram of attribute reduction

III. VARIOUS ALGORITHMS

There are various algorithms proposed for attribute reduction using various concepts of RST. Some of them are incomplete while some are complete. The time and space complexity of different algorithms are different [14].

The conventional algorithms fall into two categories:

1. Reduction algorithm based on heuristic information
2. Reduction algorithm based on random strategies

Heuristics are strategies using readily accessible, though loosely applicable information to control problem solving in human-being and machines. Fundamental heuristics is trial and error.

1. A dynamic attribute reduction algorithm based on 0-1 integer programming:

This algorithm proposed by Yitian Xu aims of making an information system into 0-1 integer table. In this algorithm he first pre-processes the constraints so that it can reduce the number of constraints and bring convenience for the solution [2]. Then a 0-1 integer programming for the original information system S_1 and its reduct R_1 is constructed. After this a 0-1 integer programming for the newly increased object and its reduct R_2 is constructed. Then finally a reduct ($R = R_1 \cup R_2$) of the dynamic information system S is achieved. The advantage of this algorithm is that it just updates the old reduct set based on new increased objects, instead of treating the change information table as new and finding the reducts. Because of this the efficiency of the

attribute reduction is improved greatly. The time and space complexity of the algorithm is $O(|C| |U|^2 / n)$ where $|C|$ denotes number of attributes, $|U|$ denotes the cardinality of the universe and $|U|/n$ denotes the number of objects in 'n' classes [2]. In this algorithm even though multiple objects enter the information simultaneously, only a few constraints are reserved after constraint pre-processing. Therefore, 0-1 programming problem can be resolved easily.

2. HeuriRed and HeuriComRed:

These two algorithms are proposed by Sango Tan, of which, HeuriRed is an incomplete algorithm. These two algorithms are categorized as Heuristic algorithms. These algorithms are coded in 'C' and tested on PC. In this type, data table is processed. In case of incomplete datasets, it is completed by conditional mean complete algorithm [2].

For conditional attributes, equally frequency binning method is adopted which is an unsupervised and univariate discretization algorithm. In this algorithm, a number 'n' is fixed as an interval and histogram of each attribute is investigated and 'n-1' cuts are determined. These 'n-1' cuts are corresponding to assignment of 'n-1' cuts such that the area between two neighboring cuts in normalized histogram is as close to '1/n' as possible [2].

Comparing Heuristic and genetic algorithms, we can say that Heuristic algorithms are faster than the genetic algorithms. Also, the attributes obtained are complete in Heuristic reduct algorithm.

The time and space complexities of both the algorithms are $O(|A||U|^2)$ where, $|A|$ denotes the number of attributes; $|U|$ denotes the number of objects of information system.

3. Attribute Reduction Based on Cross Entropy in RST:

This paper has been proposed by Jianguo Zheng and Ruixia Yan of China. In this reduction technique the authors used a different concept of cross entropy to find the reducts. Cross Entropy method can be applied to static and noisy combinational optimization problems. Entropy is an important concept to measure uncertain information. Cross entropy between two probability distribution measures the average number of bits needed to identify an event from a set of possibilities, if a coding scheme is used based on a given probability distribution Q , rather than the 'true' distribution 'P'.

$$\text{Cross Entropy, } H(P, Q) = -\sum_{i=1}^n P_i \log_2 (P_i/Q_i)$$

Where, $P = \{P_1, P_2, \dots, P_n\}$ is a probability distribution over X ,

Where $P_i \in [0, 1]$ & $\sum_{i=1}^n P_i = 1$

And $Q = \{Q_1, Q_2, \dots, Q_n\}$

Where $Q_i \in [0, 1]$ & $\sum_{i=1}^n Q_i = 1$

In this algorithm, subsets of attributes are found which has same discriminating power as original data and information and without redundancy. A greedy search can be constructed with the measures of attribute.

Two strategies are introduced, i) Forward search and ii) Backward search

In this algorithm, first the probability distribution is calculated and then cross entropy which will then help in calculating reducts.

4. An Attribute Reduction Algorithm in RST Based on Information Entropy:

This algorithm is proposed by Prof. Cui Wang and Fangfang OU of China. A new concept has been developed for this algorithm by obtaining relationship of knowledge and information theory. This algorithm was made by adopting conditional entropy and improved importance of attributes. This algorithm reduces the original attribute set by deleting the redundant and unnecessary attributes from the whole attribute set one-by-one before calculating the attribute core [11].

This algorithm is compared by MIBARK (Mutual information based algorithm from reduction of knowledge) algorithm for comparing its feasibility with others. Concept of decision table, Indiscernibility, positive region, entropy, conditional entropy and knowledge is used.

5. A New Decision Tree Algorithm Based on RST

This algorithm is proposed by Prof. Baoshi Ding of China. In this algorithm, concept of Information gain and classification is used. Compared to other algorithms, it has higher speed and accuracy. The key idea of this algorithm is to choose the attributes which have the highest information gain based on Entropy as current classification attribute and then recursively expand the branches of decision tree until the full information gain is known [10].

A tree is created by first making a node. In this, if all the samples are of same class then that node is turned into a leaf. In similar way, branches are created which in the end becomes a tree. The complexity can be calculated by the counts of nodes present in tree [10].

The programming language used in this algorithm is JAVA. Two types of concept Iterative Dichotomiser 3 (ID3) Tree and Fixed ID3 (FID3) Tree are used. By

comparing them, it was found that FID3 is better in terms of reduced complexity.

6. Matrix Algorithm for Computing Pawlak Reduction:

This algorithm is proposed by Wei Cui and Zhangyan Xu of China. Primary concern about this algorithm was that it required time consuming computation, the time complexity being $O(|C|^3|U|^2)$ [6].

To lower the time complexity of the algorithm, a new matrix was proposed and attributes reduction definition based on new matrix. Finally, significance of the attribute was defined and the time complexity is cut to $O(|C||U|) + O(|C|^2|U/C|^2)$ [6].

For improving the efficiency of attribute reduction, new heuristic information, with new Boolean matrix was designed. The space complexity is $O(|U|^2)$.

7. An Efficient Decision Tree Algorithm Based on Concentrated Ordered Discernibility Set:

This algorithm is proposed by Prof. Chuanjioi Yang and Hao Cje of China. The algorithm is based on the concept of discernibility matrix needs to occupy a lot of memory space and many elements of discernibility matrix are not necessary for reduction. This algorithm need not to create discernibility matrix and generate redundant elements, so it can cut down the storing and the computing capacity greatly and the efficiency of algorithm is improved. The time and space complexity of this algorithm is $O(|C|^2|U|^2)$. The basic idea is as follows: at first, creating discernibility matrix and then selecting the attribute based on heuristics and deleting the elements which contain the attribute and so on, until discernibility matrix is empty. The concept of algorithm based on positive region and reduction algorithm based on concentrated ordered discernibility set is used. The space complexity of the positive region algorithm is smaller than the reduction algorithm based on discernibility matrix but the time complexity is almost similar [9].

8. A Decision Tree Algorithm Based on RST after Dimensionality Reduction

The algorithm is proposed by Indian Prof. Shailendra K. Shrivastava and Manisha Tantuway. In this algorithm firstly large volume of dataset which contain redundant instance are reduced. These redundant instances doesn't make any contribution to take decision hence can be deleted from the dataset. After reducing the volume of the dataset decision tree is constructed through rough set [3]. The main concept of rough set theory in this algorithm is degree of dependency which is used in the proposed algorithm to select splitting attribute on the compressed data. Thus the proposed algorithm reduces the complexity of tree and in addition

increases the accuracy. Firstly in this algorithm duplicate values are eliminated and thus the size of the dataset is reduced. By the deletion of duplicate instance the degree of dependency of decision attribute on all condition attribute is calculated. If dependency of decision attribute on condition attribute is zero then this condition attribute is irrelevant and doesn't make any contribution to take decision. After calculating degree of dependency of decision attribute on condition attribute, the condition attribute on which degree of dependency of decision attribute has maximum is selected as splitting attribute. This process is repeated in every level until all samples are classified [3]. Then the splitting attributes are branched having different attribute values and a decision tree is made by repeating this process. The proposed algorithm has also contributed to reduce noise and irrelevant information from the tree. Compared to ID3 algorithm this algorithm gives precise and the dimensionality of the database is also reduced.

COMPARISON OF ATTRIBUTE REDUCTION ALGORITHMS ON THE BASIS OF TIME AND SPACE COMPLEXITY

Name of algorithm	Time complexity	Space complexity
Dynamic algorithm based on 0-1 integer programming	$O(A U ^2/n)$	$O(A U ^2/n)$
Johnson approximation algorithm	$O(A ^2 U ^2)$	$O(A U ^2)$
Feature ranking mechanism by Keyun Hu	$O((A + \log U) U ^2)$	$O(A U ^2)$
Vinterbo algorithm	$O(A ^2 U \log U)$	$O(U)$
Jue Wang algorithm	$O(A ^2 U ^2)$	$O(A U ^2)$
HeuriRed algorithm (Incomplete algorithm)	$O(A U ^2)$	$O(A U ^2)$
HeuriComRed algorithm	$O(A U ^2)$	$O(A U ^2)$
Matrix Algorithm for computing Pawlak reduction	$O(C U) + O(C ^2 U/C ^2)$	$O(U ^2)$

Table 3: Comparison of attribute reduction algorithms

IV. SIMULATION RESULTS

We implemented Quick Reduct algorithm. This algorithm uses various concepts of rough set theory like dependency, positive region, and equivalence class. We worked on a medical database and were able to reduce four attribute out of nine conditional attributes.

The evaluation Quick Reduct algorithm can be done using following steps:

1. Scan the database.
2. Find the equivalence class of conditional and decision attributes.
3. Find the subset of each conditional with the decision class.
4. If the dependency is 1 then eliminate the conditional attribute.
5. Else go to step 3.
6. Compare all conditional attributes and get the reduced database.

Using this algorithm we got the following simulation results for the following database. This database consists of 9 features of 9 persons. These nine features (redness, swelling, Watery, photophobia etc.) are conditional attributes based on which decision is made i.e. whether patient have conjunctivitis or not.

U	Redness	Swelling	Watery	Photophobia	Itching	Headache	Pain	Discharge	Loss of vision	Conjunctivitis
1	1	0	0	0	1	0	0	1	0	0
2	0	0	1	1	1	1	0	1	0	0
3	1	0	1	1	0	1	1	1	1	1
4	1	1	0	0	1	1	0	0	1	1
5	0	0	1	1	0	1	1	1	0	1
6	1	0	0	0	1	0	1	1	0	0
7	0	1	0	0	1	0	0	0	1	1
8	0	1	0	0	0	1	1	1	0	0
9	1	1	1	1	1	1	1	1	1	1

Table 4: Medical Database

Applying steps of quick reduct algorithm on medical database gives reduct as Redness, Swelling, Photophobia, Headache and discharge.

Quick reduct algorithm is basically dependent on relative dependency of the sets. The efficient reduction of attributes is achieved by comparing equivalence relations calculated by sets of attributes. Attributes are removed so that the reduced set provides the same predictive capability of the decision feature as the

original. If the relative dependency between two sets is 1, then one of the set is removed, confirming that no information is lost by removing the set. In this way the database is reduced by finding the reducts.

The design was implemented using Xilinx software and functional verification was done. The simulation results are shown in Fig.1.

V. CONCLUSIONS AND FUTURE WORK

In this paper we presented various algorithms for attribute reduction using various concepts of rough set theory. Also we compared them with their space and time complexity. We also listed various types of algorithms for the same. The implementation of an algorithm for a generic database and comparing the same algorithm with others is our future work.

VI. REFERENCES

- [1] Yueping Xian, Liyu Huang, JingZhang Liang, Ying Pan, A Complete Attribute Reduction Algorithm Based on Improved FP Tree, 2011 IEEE.
- [2] Yitian Xu , Laisheng Wang, Ruiyan Zhang, A dynamic attribute reduction algorithm based on 0-1 integer programming, Published by Elsevier B.V. in 2011 in Science Direct.
- [3] Shailendra K. Shrivastava, Manisha Tantuway, A Decision Tree Algorithm based on Rough Set Theory after Dimensionality Reduction, International Journal of Computer Applications Volume 17– No.7, March 2011.
- [4] C. Velayutham and K. Thangavel, Unsupervised Quick Reduct Algorithm Using Rough Set Theory, JOURNAL OF ELECTRONIC SCIENCE AND TECHNOLOGY, VOL. 9, NO. 3, SEPTEMBER 2011.
- [5] P.Kalyani, Dr. M.karnan, A new implementation of Attribute reduction using Quick Relative Reduct algorithm, International Journal of Internet Computing, Volume-I, Issue-1, 2011
- [6] Zhangyan Xu, Wei Cui, Matrix Algorithm for Computing Pawlak Reduction, IEEE 2009.
- [7] Jiang Yu, Du Bin, An Efficient Complete Algorithm for Attribute Reduction, National Natural Science Foundation of China, 2009 IEEE.
- [8] Cuiru WANG, Fangfang OU, An Attribute Reduction Algorithm Based on Conditional Entropy and Frequency of Attributes, National

- Natural Science Foundation of China, IEEE 2009.
- [9] Qingguo Zhang, Xuefeng Zheng, Zhangyan Xu, Efficient Attribute Reduction Algorithm Based on Incomplete Decision Table, 2009 Second International Conference on Intelligent Computation Technology and Automation, IEEE Computer society.
- [10] Baoshi Ding, Yongqing Zheng, Shaoyu Zang, A New Decision Tree Algorithm Based on Rough Set Theory, 2009 Asia-Pacific Conference on Information Processing.
- [11] Cuiru WANG, Fangfang OU, An Attribute Reduction Algorithm in Rough Set Theory Based on Information Entropy, 2008 IEEE.
- [12] Cuiru WANG, Fangfang OU, an Attribute Reduction Algorithm in Rough Set Theory Based on Information Entropy, 2008 IEEE.
- [13] Weiwei Fang, Bingru Yang, Wei Hou, Changsheng, Zhou, New heuristic attribute reduction algorithm based on Rough set theory, IEEE 2008, Beijing Information Science and technology University 2008.
- [14] Hongbo Xu, Jun Gu- Hong Kong University of Science and Technology, Hong Kong , Efficient algorithms for attributes reduction problems by Songbo Tan , International Journal of Innovative Computing, Information and Control Volume 1, Number 4, December 2005
- [15] Z Suraj, An Introduction to Rough Set Theory and Its Applications, December 27-30, 2004, Cairo, Egypt.
- [16] Rough set theory and its applications by Zdzisław Pawlak, Journal of telecommunications and Technology, 3/2002.
- [17] Pawlak, Z.: Rough Sets – Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht Boston London, (1991).

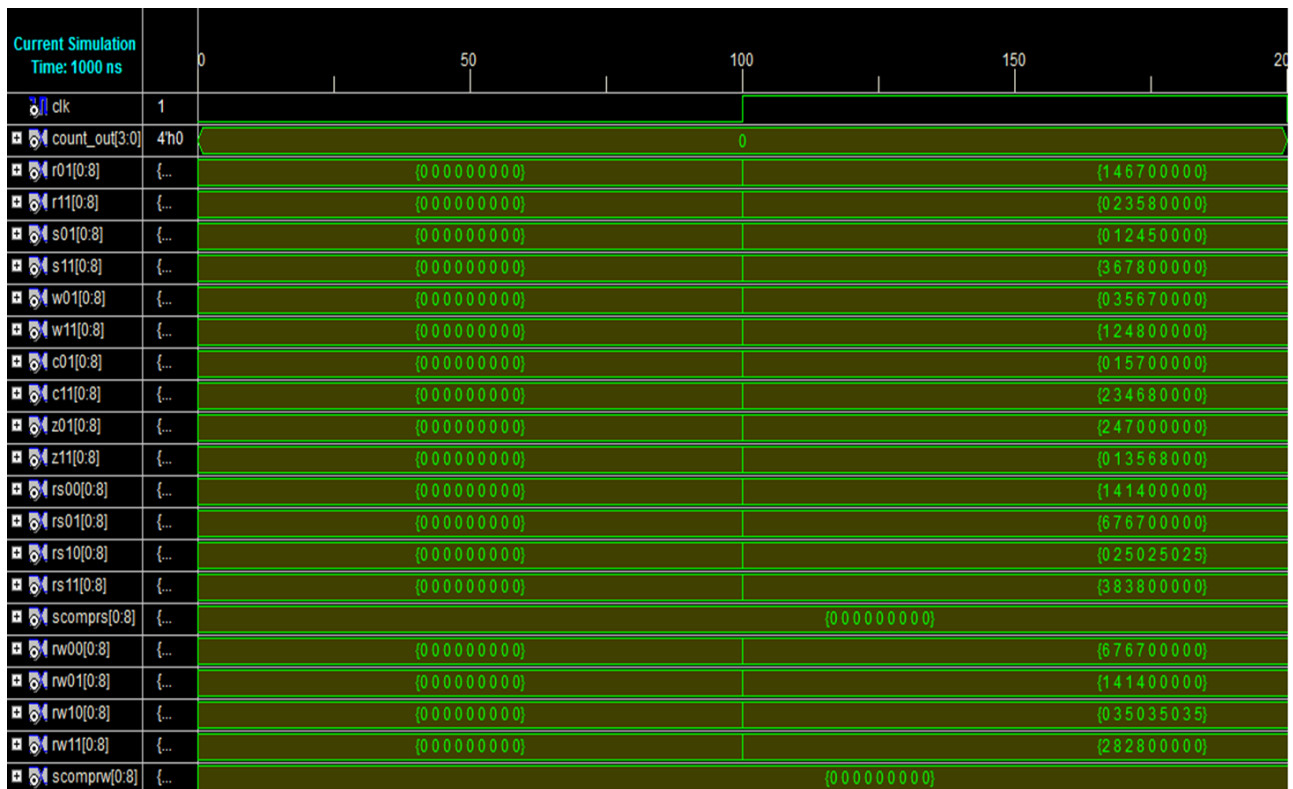


Fig. 1: Simulation results using Quick Reduct algorithm (1)

