Classification of n-Variable Boolean Functions through Hamming Distance and their Application in System Biology

Jayanta Kumar Das
Govt. College of Engineering and Ceramic Technology
73,A.C.Banerjee Lane,Kolkata-700010

Pabitra Pal Choudhury
Applied Statistics Unit
Indian Statistical Institute
203, B. T. Road, Kolkata-700108.

ABSTRACT
In this article, a noble approach is presented to classify n-variable Boolean functions in logical way such that each function belonging to a particular class can be traced with respect to a single base function. In the present study, two different methods have been proposed for this classification. The first one is done through the Hamming distance with regards to base 0 (2^n bits of zeros) Boolean function. In the second method, the classification is done to generate all Boolean functions from n variable to n+1 variable through the concatenation methodology. The presented paper also contains two unique and different methodologies for finding the cardinality of different classes. In this classification all the basis Boolean functions were captured into a single class. All the linear and corresponding affine Boolean functions belong to a single class along with other nonlinear Boolean functions except two classes of single cardinality. It has been also observed symmetrical class distribution with equal cardinality and functions belonging to the symmetrical classes are complement of each other. Special Boolean functions like Nested Canalyzing Functions (NCFs) [1, 2, 3, 4, and 5] are considered biologically important. So they are specially viewed in our classification among different classes.

Keywords
Classification Methodology; Boolean function; Hamming Distance; Nested Canalyzing Function; Interaction Graph.

1. INTRODUCTION
Boolean functions have many applications in system biology as well as in the theory of computer science. But only few set of Boolean functions have well replication in actual behavior of nature. Classification of n-variable Boolean functions is a challenging task with the requirement of uniform distribution of Boolean functions among different classes. It is very much vital and meaningful to classify all those Boolean functions which are biologically important. Some biological important function such as Nested Canalyzing Functions (NCFs) is best suited to reduce the chaotic behavior and to attain stability in gene regulatory network. NCFs and its variants have wide range of application in system biology. Further Interaction graph [6, 7] is used for static analysis of Boolean network. It has been observe some NCFs with fixed hamming distance are capable of forming the similar Interaction Graphs in which all the vertices are connected and their symmetric classes also form the similar Interaction graphs with equal positive and negative edges.

In [8], the work has been done for the classification of n-variable Boolean functions where affine functions are uniformly distributed and each affine function are capable to capture all others Boolean functions uniformly. Earlier, when two Boolean functions of n-variable differ only by permutation or complementation of their variables, they fall into equivalence classes. The formula for counting the number of such equivalence classes is given in [9]. Further, it has also been elaborated in [10] about the procedures of selection of a representative assembly, with one member from each equivalence class. In [11], the linear group and the affine Boolean function group of transformations have been defined and an algorithm has been proposed for counting the number of classes under both groups. The classification of the set of n-input functions is specifically based on three criteria: the number of functions, the number of classes, and the number of NPN classes, which were first introduced in [12]. In [13] the work has been done for classification of Boolean functions where each functions belonging to the equivalence classes differ by only permutation and complementation of their variables. Three Variable Boolean functions in the name of 3-neighbourhood Cellular Automata rules are classified on the basis of Hamming distance with respect to linear rules **[**14**]**. The characterization of 3-variable non-linear Boolean functions in the name of 1-D cellular automata rules is under taken in three different ways by Boolean derivatives [15], by deviant states [14] and by matrices [16]. Classification of the affine equivalence classes of cosets of the first order Reed-Muller code with respect to cryptographic properties such as correlation immunity, resiliency and propagation characteristics are discussed in [17]

In section 2, classification methodology has been discussed and in section 3, cardinality of the classes was found. Section 4, provides results and application of the classes.

2. CLASSIFICATION METHODOLOGY
An n-variable Boolean function $f(x_1, x_1, x_1, \dots, x_1)$ is a mapping$\{0,1\}^n \rightarrow \{0,1\}$. The Number of different n-variable Boolean functions is 2^{2^n} where each function can be represented by a truth table output as a binary string of length 2^n.

Hamming Distance (H.D) between two Boolean functions is denoted as $H.D(f, g) = k$, where k can be

$0, 1, 2, \ldots, 2^n - (n+1)$ and f and g are two Boolean functions, one is consider as base function. In this section two different classification methodologies have been proposed to classify the set of all possible n-variable Boolean functions with the help of Hamming Distance. One is through the direct hamming distance and another is from n variable to n+1 variable with the help of concatenation procedure.

2.1 Through the Direct Hamming Distance

The proposed method for classification of n-variable Boolean functions is through the hamming distance with regards to base Boolean function. First, effort has been given to find all possible Hamming distances with regards to base Boolean function and the number of possible functions for any particular hamming distance. As the length of the string for n variable is 2^n bits, so number of possible Hamming distance for n variable are 0, 1, 2… 2^n . Thus the number of classes of n variable is given by $2^n + 1$. As 2^n is even for all positive integer n, so number of class will be odd. Finding the cardinality of each class is explained in section 3.

Example 1: For n=2, there are $2^2 + 1 = 5$ classes denoted by **Class-I**, where 0<=I<=2^n are shown in Table I.

Table 1. Number of classes and functions distribution into different classes for n=2 variable

Class-0	Class-1	Class-2	Class-3	Class-4
0000	0001 0010 0100 1000	1100 0011 1010 0101 1001 0110	1110 1101 1011 0111	1111

2.2 Through the Concatenation Procedure from n variable to n+1 variable

Let, A=$\{ab\}$ and B=$\begin{Bmatrix} c \\ d \end{Bmatrix}$ are two sets; concatenation of these two sets in order f_1 =A×B=$\begin{Bmatrix} abc \\ abd \end{Bmatrix}$; f_2 =B×A=$\begin{Bmatrix} cab \\ dab \end{Bmatrix}$ and addition of $f_1$and $f_2$defined as f_1+f_1=$\begin{Bmatrix} abc \\ abd \\ cab \\ dab \end{Bmatrix}$.

Suppose for n=1 variable, there are 3 classes with 0, 1 and2 possible hamming distances are ${C^1}_0$=$\{00\}$, ${C^1}_1 = \begin{Bmatrix} 01 \\ 10 \end{Bmatrix}$ and ${C^1}_2$=$\{11\}$; ($C_k \cong Class - k$);

Now for n=2 variable, 5 classes are present. To find all the classes of 2 variables concatenation procedure has been used as follows:

${C^{n+1}}_k$=$\sum_{i=0}^{2^n} {C^n}_i \times {C^n}_j$; where k=i + j and i≤ 2^n; j≤ 2^n; 0≤ $k \leq 2^{n+1}$;

${C^2}_0$=$\{C_0 \times C_0\} = \{00\} \times \{00\} = \{0000\}$

${C^2}_1$=$\{C_0 \times C_1\} + \{C_1 \times C_0\} = \{00\} \times \begin{Bmatrix} 01 \\ 10 \end{Bmatrix} + \begin{Bmatrix} 01 \\ 10 \end{Bmatrix} \times \{00\} = \begin{Bmatrix} 0001 \\ 0010 \end{Bmatrix} + \begin{Bmatrix} 0100 \\ 1000 \end{Bmatrix}$=$\begin{Bmatrix} 0001 \\ 0010 \\ 0100 \\ 1000 \end{Bmatrix}$

${C^2}_2$=$\{C_0 \times C_2\} + \{C_1 \times C_1\} + \{C_2 \times C_0\} = \{00\} \times \{11\} + \begin{Bmatrix} 01 \\ 10 \end{Bmatrix} \times \begin{Bmatrix} 01 \\ 10 \end{Bmatrix} + \{11\} \times \{00\} = \{0011\} + \begin{Bmatrix} 0101 \\ 0110 \\ 1001 \\ 1010 \end{Bmatrix} + \{1100\}$=$\begin{Bmatrix} 0011 \\ 1100 \\ 0101 \\ 0110 \\ 1001 \\ 1010 \end{Bmatrix}$

${C^2}_3$=$\{C_1 \times C_2\} + \{C_2 \times C_1\} = \begin{Bmatrix} 01 \\ 10 \end{Bmatrix} \times \{11\} + \{11\} \times \begin{Bmatrix} 01 \\ 10 \end{Bmatrix} = \begin{Bmatrix} 0111 \\ 1011 \end{Bmatrix} + \begin{Bmatrix} 1101 \\ 1110 \end{Bmatrix}$=$\begin{Bmatrix} 0111 \\ 1011 \\ 1101 \\ 1110 \end{Bmatrix}$

${C^2}_4$=$\{C_2 \times C_2\} = \{11\} \times \{11\} = \{1111\}$.

So using concatenation methodology from n=1 variable, all the classes ($Class - 0,\ Class - 1, Class - 2, Class - 3, Class - 4$) for n=2 variable is found.

3. FINDING CARDINALITY OF DIFFERENT CLASSES

Cardinality of each class can be obtained in two ways- one is through hamming distance for n variable and another is for finding the cardinality for n+1 variable if all the cardinalities of n variable are known.

3.1 Through the Hamming Distance

Let, m be the Hamming Distance for n variable, then the cardinality of the class m denoted by **Class-** $\boldsymbol{m = 2^n{}_{C_m}}$; where m<=$2^n$.

Example 2: Suppose, to find the cardinality of the **Class-m** for n variable; denoted by CD_m^n, using the formula the cardinality obtained is

$CD_2^2(n = 2, m = 2) = 4_{C_2}$=6**.** This is exactly same as shown in Table 1 for **Class-2** and for **Class-3** 4_{C_3}=4 so on.

$$C_{2=}\begin{Bmatrix} 1100 \\ 0011 \\ 1010 \\ 0101 \\ 1001 \\ 0110 \end{Bmatrix} = 6\ , \quad C_{3=}\begin{Bmatrix} 1110 \\ 1101 \\ 1011 \\ 0111 \end{Bmatrix} = 4$$

3.2 Finding the Cardinality of Class-k for n+1 variable

If $x_0, x_1, x_2, \ldots, x_m$ denote the cardinalities of all the classes for n variable, then cardinality (CD_k^{n+1}) of the Class-k for n+1 variable can be given by the formula as:

CD_k^{n+1}= $\sum_{i=0}^{2^{n+1}} G(i,j)$ where i+j=k; 0≤k≤2^{n+1}; 0≤m≤2^n and

$$G(i,j) = \begin{Bmatrix} x_i \times x_j \;\; if\ i,j \leq 2^n; \\ 0\ if\ i\ or\ j > 2^n; \end{Bmatrix}$$

Example 3: Suppose, to find the cardinality of the **Class-k (k=2)** for n=3 variable**, a**ll the cardinalities of each **Class** in n=2 variable has to be known. All the cardinalities for 2 variable **Classes** is obtained from Table 1 as $CD_{0=}1$,

CD_1 =4, $CD_2 = 6$, $CD_3 = 4$ and $CD_4 = 1$, here m=4. $(CD_i \cong x_i)$

So CD_k^{n+1} $(k = 2, n = 2) =$ $CD_2^3 = G(0,2) + G(1,1) + G(2,0)$

= (1×6) + (4×4) + (6×1)

= 6+16+6

= 28

Now using the formula **Class** $\boldsymbol{m = 2^n{}_{C_m}}$ described in section 3.1 the cardinality of Class-**2** is $C_2 = \boldsymbol{2^3{}_{C_2}} =$ $\boldsymbol{8_{C_2} = 28}$ **which is** same as previous formula.

4. RESULTS

4.1 Symmetric relations between Classes

Let, $\boldsymbol{f}$ is the set of functions for **Class-J**, then the symmetric class of **Class-J** is **Class-($\boldsymbol{2^n}$-j)** which contains the set of functions $\boldsymbol{f'}$**,** complement of $\boldsymbol{f}$ and vice versa; where 0≤j≤$2^n - j$; Number of symmetric relation among $2^{n-1} + 1$ pairs are shown in Table 2. Now, observed that all the basis functions belong to Class-1 for n variable and number of basis functions $= \boldsymbol{2^n}$ i.e. exactly same as cardinality of the Class-1 ($\boldsymbol{2^n{}_{C_1}}$).

Table 2 Symmetrical Class distribution

Variable (n)	**Number of Classes for n variable**								
1				**1**	**2**	**1**			
2			**1**	**4**	**6**	**4**	**1**		
3	**1**	**8**	**28**	**56**	**70**	**56**	**28**	**8**	**1**

Example 4: For n=2, there are 5 classes and there are 3 pairs of class that forms the symmetric relations. Middle class always formed the symmetric relation w.r.t itself i.e. if any function belongs to the middle class, then the complement of the function also belongs to the middle class.

Class-0~Class-4, Class-1~Class-3, Class-2~Class-2;

$Class - 0 = \{0000\} \sim Class - 4 = \{\mathbf{1111}\}$;

$$Class - 1 = \begin{Bmatrix} 0001 \\ 0010 \\ 0100 \\ 1000 \end{Bmatrix} \sim Class - 3 = \begin{Bmatrix} \mathbf{1110} \\ \mathbf{1101} \\ \mathbf{1011} \\ \mathbf{0111} \end{Bmatrix};$$

$$Class - 2 \sim Class - 2 = \begin{Bmatrix} 1100 \\ \mathbf{0011} \\ 1010 \\ \mathbf{0101} \\ 1001 \\ \mathbf{0110} \end{Bmatrix}$$

4.2 Identification of NCFs through the classes

All the NCFs [2] are always odd hamming distance with respect to base function, and we have all possible (1, 3, 5 …$2^n - 1$) odd hamming distances classes. The first and last odd class (1 and $2^n - 1$) consist of NCFs and others are distributed symmetrically into different odd classes.

Example 5: For n=3, there are 64 [2] nested canalizing functions and for n=3, there are 1, 3, 5 and 7 odd classes. Cardinality of first and last odd class is 8, so both these two classes contain 16 NCFS; other 48 distributed in Class-3 and Class-5 symmetrically each having 24 Boolean functions.

4.3 Interaction graph for different classes

The Interaction graph is denoted by $G(f)$=(V, E), is the sign directed graph on vertex set V∊$\{1,2,3 \ldots n\}$ corresponds to node and edge set E ∊$\{+,-\}$, an arc (positive or negative) between nodes.

Interaction Graph is formed using Boolean function and number of nodes is equal to number functions. It has been observed that if n functions with hamming distance 1 (i.e. for **Classe-1)** is taken, then the graph are complete connected graph. In the context of gene regulatory network it can be said that there is a path from one component to the rest. All the functions belonging to a particular **Class** and the corresponding symmetric **Class** are capable of forming equivalent Interaction Graph. In other words, equivalent classes form isomorphic Interaction Graph.

Example 6: For n=3, suppose three Boolean functions has been taken as f_1=1, f_2=4, and f_3=8, from **Class-1** and Interaction Graph is shown in Figure 1 in which there exist a path from one node to other. In Figure 2 Interaction Graph for complement of those three functions as $\bar{f}_1$=127, $\bar{f}_2$=251and $\bar{f}_3$=247 from **Class-7** there is exist a path from one node to other. It has been seen that both the graphs have equal positive and negative signs and equal number of cycles.

Figure 1. Interaction Graph for the functions $\boldsymbol{f_1}$, $\boldsymbol{f_4}$ and $\boldsymbol{f_8}$

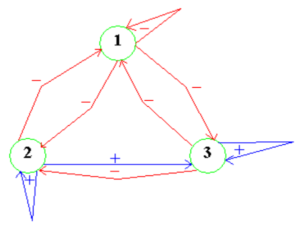
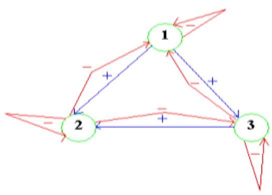


Figure 2. Interaction Graph for the functions $\boldsymbol{f_{127}, f_{251}}$ and $\boldsymbol{f_{247}}$



5. REFERENCES

[1] L. Layne, E.Dimaitrova, M. Macauley, " Nested Canalyzing Depth and Network Stability",Bull Math Biol (2012) 74:422433 DOI 10.1007/s11538-011-9692-y

[2] Camellia. Ray, Jayanta Kumar das, Pabitra Pal Choudhury, “On Analysis and Generation of some Biologically Important Boolean Functions” arXiv:submit/0952437 [cs.YS] 8 Apr 2014 .

[3] Abdul Salam Jarrah, Blessilda Raposa, Reinhard Laubenbacher " Nested Canalyzing Unate Cas- cade,

and Polynomial functions Physica D. 2007 September 15; 233(2): 167174

[4] Franziska Hinkelmann, Abdul Salam Jarrah, " Inferring Biologically Relevant Models: Nested Canalyzing Functions ", International Scholarly Research Network ISRN Biomathematics Volume 2012, Article ID 613174, 7 pages doi:10.5402/2012/613174

[5] David Murrugarra and Reinhard Laubenbacher, "The Number of Multistate Nested Canalyzing Functions ", Physica D: Nonlinear Phenomena Volume 241, Issue 10, 15 May 2012, Pages 929938

[6] L. Pauleve, A. Richard, "Stastic Analysis of Boolean Networks Based on Interaction graph: A survaey" ,Theoretical Computer scinece, 284(2012):93-104.

[7] Ranjeet Kumar Rout, Jayanta Kumar Das, Pabitra Pal Choudhury, "Analysis of Boolean Functions based on Interaction Graphs and their influence in System Biology" (Accepted for seminar and possible publication in your upcoming symposium "3rd International Symposium on Complex Dynamical Systems and Applications (CDSA -2014).

[8] R. K. Rout, P. Pal Choudhury, and S. Sahoo, "*Classification of Boolean Functions Where Affine Functions Are Uniformly Distributed,*" *Journal of Discrete Mathematics*, vol. 2013, Article ID 270424, 12 pages, 2013.

[9] D. Slepian,"On the number of symmetry types of Boolean functions of n variables" Canadian Journal of Mathematics, vol. 5, no. 2, pp. 185–193, 1953. View at Zentralblatt MATH · View at MathSciNet.

[10] S. W. Golomb, "On the classification of Boolean functions," IRE Transactions on Circuit Theory, vol. 6, no. 5, pp. 176–186, 1959.

[11] M. A. Harrison, "On the classification of Boolean functions by the general linear and affine groups" Journal of the Society for Industrial and Applied Mathematics, vol. 12, no. 2, pp. 285–299, 1964.

[12] V. P. Correia and A. I. Reis, "Classifying n-input Boolean functions," in Proceedings of the 7th Workshop IBERCHIP (IWS '01), pp. 58–66, Montevideo, Uruguay, March 2001.

[13] S. W Golomb (1959), " On the classification of Boolean functions", *IRE transactions on circuit theory, Vol. 06, Issue. 05, pp.176- 186* .

[14] P. Pal Choudhury, S. Sahoo, M. Chakarborty, S. K Bhandari, A. Pal(2009), "Investigation of the Global Dynamics of Cellular Automata Using Boolean Derivatives ", Computers and Mathematics with Applications, *Elsevier, 57, pp. 1337-1351* .

[15] P. Pal Choudhury, S. Sahoo, M. Chakraborty (2011) ,"Characterization of the evolution of Non-linear Uniform Cellular Automata in the light of Deviant States", *Mathematics and Mathematical Sciences*, Vol. 2011, Article ID 605098.

[16] S. Sahoo, P. Pal Choudhury, M. Chakraborty (2010), "Characterization of any Non-linear Boolean function Using A Set of Linear Operators", *Journal of Orissa Mathematical Society, Vol. 2, No. 1&2, pp. 111-133.*

[17] An. Braeken, Y. Borissov, S. Nikova, B. Preneel (2005), "Classification of Boolean Functions of 6 Variables or Less with Respect to Cryptographic Properties", ICALP'05 Proceedings of the 32nd international conference on Automata, Languages and Programming,Springer-Verlag Berlin, Heidelberg , pp. 324-334.