Fitness Landscape and Evolutionary Boolean Synthesis using Information Theory Concepts

Arturo Hernández Aguirre
Center for Research in Mathematics, Department of Computer Science
Mineral de Valenciana, Guanajuato, 36240, México
artha@cimat.mx

Carlos Coello Coello
CINVESTAV-IPN Computer Science Section
México, D.F. 07300, México
ccoello@cs.cinvestav.mx

Abstract

In this paper we show how information theory concepts can be used in evolutionary circuit design and minimization problems. Conditional entropy, mutual information, and normalized mutual information are commonly used to measure or estimate the amount of information shared by two random variables. Although the simple number reported by these measures may guide the evolutionary search, we show that normalized mutual information produces more amenable fitness landscape for search than the others. Several landscape plots and experiments are used to support and explain our main argument.

1 Introduction

In this paper we use multiplexers and genetic programming (GP) for the synthesis of gate-level Boolean functions. This means that GP working at a gate-level representation will try to produce the circuit that implements the Boolean function. We propose a fitness function that by measuring the Normalized Mutual Information, drives the search towards circuits that maximize the similarity between the target function and the evolved function.

2 Problem Statement

The design problem is the following: find the smallest circuit that implements a Boolean function specified by its truth table [2, 1]. The design metric adopted in this case is the number of components in a 100% functional circuit. The process works at "gate-level" and the only component replicated is the binary multiplexer.

3 Basic concepts of IT

Uncertainty and its measure provide the basis for developing ideas about Information Theory [3]. The most commonly used measure of information is Shannon's entropy.

Entropy The average information supplied by a set of k symbols whose probabilities are given by $\{p_1, p_2, \ldots, p_k\}$, can be expressed as,

$$H(p_1, p_2, \ldots, p_k) = -\sum_{s=1}^{k} p_k log_2 p_k \qquad (1)$$

The information shared between the transmitter and the receiver at either end of the communication channel is estimated by its Mutual Information,

$$MI(T;R) = H(T)+H(R)-H(T,R) = H(T)-H(T|R) \qquad (2)$$

The conditional entropy $H(T|R)$ can be calculated through the joint probability, as follows:

$$H(T|R) = -\sum_{i=1}^{n}\sum_{j=1}^{n} p(t_i r_j) log_2 \frac{p(t_i r_j)}{p(r_j)} \qquad (3)$$

An alternative expression of mutual information is

$$MI(T;R) = \sum_{t\in T}\sum_{r\in R} p(t,r) log_2 \frac{p(t,r)}{p(t)p(r)} \qquad (4)$$

Mutual information, Equation 2, is the difference between the marginal entropies $H(T) + H(R)$, and the joint entropy $H(T, R)$. We can explain it as a measure of the amount of information one random variable contains about another random variable, thus it is the reduction in the uncertainty of one random variable due to the knowledge of the other.

Studholme [5] proposed normalized mutual information as an invariant measure for image registration problems. His approach improves mutual information since he shows his normalized version has better characteristics for measuring the shared information between two images at different angles and area of overlapping. We investigate this issue in Section 4

$$NMI(T; R) = \frac{H(T) + H(R)}{H(T, R)} \tag{5}$$

4 Entropy and Circuits

Entropy has to be carefuly applied to the synthesis of Boolean functions. Assume any two Boolean functions, $F1$ and $F2$, and a third $F3$ which is the one's complement of F2, then $F3 \neq F2$.

$$H(F2) = H(F3)$$

Also Mutual Information shows a similar behaviour.

$$MI(F1, F2) == MI(F1, F3)$$

Assume the target Boolean function is T, then $MI(T, F2) = MI(T, F3)$, but only one of the circuits implementing F2 and F3 would evolve towards the solution since their Boolean functions are complementary. A fitness function based on mutual information will reward both circuits with the same value, but their Boolean function could be complementary. When these individuals are selected for reproduction they destroy their building blocks hence convergence is never reached. The events just described are similar to the TwoMax problem and the Ising Model problem of Hoyweghen et.al. [4]. We propose the use of Hamming distance as a way to bias the search towards only one of the attractors (see next section).

$$badfitnessfunction\#1 = MI(T, C) = H(T) - H(T|C)$$

The entropy term $H(T)$ is constant since this is the expected target vector. Therefore, instead of maximizing mutual information the fitness function can only minimize the conditional entropy,

$$badfitnessfunction\#2 = H(T|C) \tag{6}$$

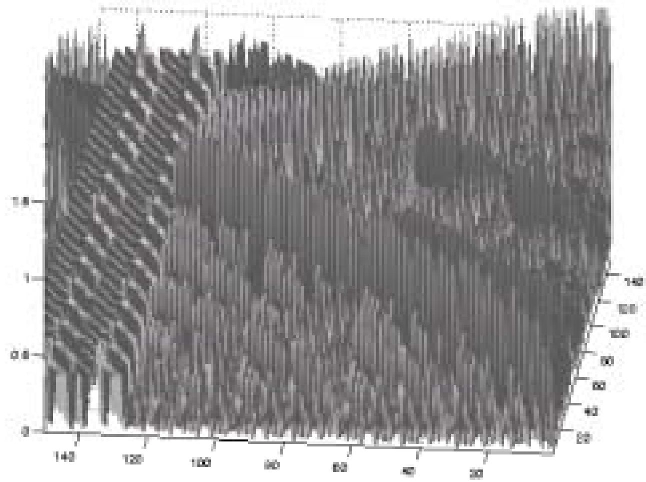

Figure 1. Landscape of Mutual Information

We called *bad* to these fitness functions based on mutual information because we were not able to find a solution with them. Although mutual information has been described as the "common" information shared by two random processes, the search space is not amenable for evolutionary computation. In Figure 1 we show this search space over mutual information for all possible combinations with two binary strings of 8 bits (shown in decimal). The area shown corresponds to about $\frac{1}{4}$ ($[1, 150] \times [1, 150]$) of the whole search space of ($[1, 254] \times [1, 254]$) (the values 0 and 255 were not used).

For any two equal vectors, their Mutual Information lies on the line at 45^o (over points $\{(1, 1), (2, 2), (3, 3) \ldots (n, n)\}$). In the next Section we continue this discussion and design fitness functions whose *landscape* seems more promisory for exploration.

5 Fitness Function based on Normalized Mutual Information

So far we have described the poor scenario where the search is driven by a fitness function based on the sole mutual information. We claim that fitness functions based on Normalized Mutual Information (NMI) should improve the performance of the genetic programming algorithm because of the form of the NMI landscape. This is shown in Figure 3 for two 8-bit vectors (as previous case). Note on the figure how the search space becomes more regular, and more important, notice the appearance of the *wall* at 45^o where both strings are equal.

We propose three new fitness functions based on Normalized Mutual Information (Equation 5) and report experiments using the next three fitness functions (higher fitness means better).

Assume a target Boolean function of m atributes $T(A_1, A_2, \ldots, A_m)$, and the circuit Boolean function C of the same size. In the following, we propose variations of the basic fitness function of Equation 7, and discuss the intuitive idea of their (expected) behavior.

$$fitness = (Length(T) - Hamming(T,C)) \times NMI(T,C) \tag{7}$$

We tested Equation 7 in the synthesis of several problems and the results were quite promising. Thus, based on this primary equation we designed the following fitness functions. In Figure 2 we show the *fitness landscape* of Equation 7.

$$fitness1 = \sum_{i=1}^{m} \frac{fitness}{NMI(A_i, C)} \tag{8}$$

$$fitness2 = \sum_{i=1}^{m} fitness \times NMI(A_i, C) \tag{9}$$

$$fitness3 = (Length(T) - Hamming(T,C)) \times (10 - H(T|C)) \tag{10}$$

The function fitness, Equation 7, is driven by NMI(T,C) and adjusted by the factor $Length(T) - Hamming(T,C)$. This factor tends to zero when T and C are far in Hamming distance, and tends to $Length(T)$ when T and C are close in Hamming distance. The effect of the term is to give the correct rewarding of the NMI to a circuit C close to T. Equation 7 is designed to remove the convergence problems described in the previous section.Fitness1 and Fitness2, Equations 8 and 9, combines NMI of T and C with NMI of C and the attributes A_k of the target function. Thus, fitness1 and fitness2 pretends to use more information available in the truth table in order to guide the search. Fitness3 is based on conditional entropy and it uses the mentioned factor to supress the reproduction of undesirable trees. Since conditional entropy has to be minimized we use the factor $10 - H(T|C)$ in order to maximize fitness. Equations 8 and 6 use the conditional entropy term, nevertheless, only Equation 8 works fine. As a preliminar discussion regarding the design of the fitness function, the noticeable difference is the use of Hamming distance to guide the search towards the aforementioned *optimum wall* of the search space. The Hamming distance destroys elements of the population on one side of the wall, and favors the other side. Thus, there is only one attractor in the search space.

6 Experiments

In the following experiments we find and contrast the convergence of our GP system for the three fitness functions defined above.

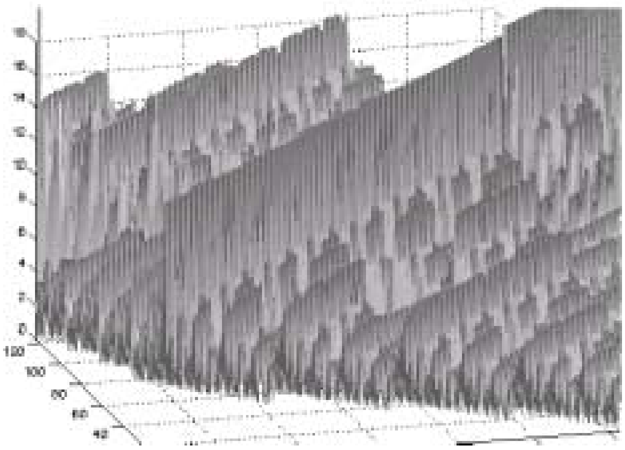

Figure 2. Landscape of: $f = (Length(T) - Hamming(T,C)) \times NMI(T,C)$

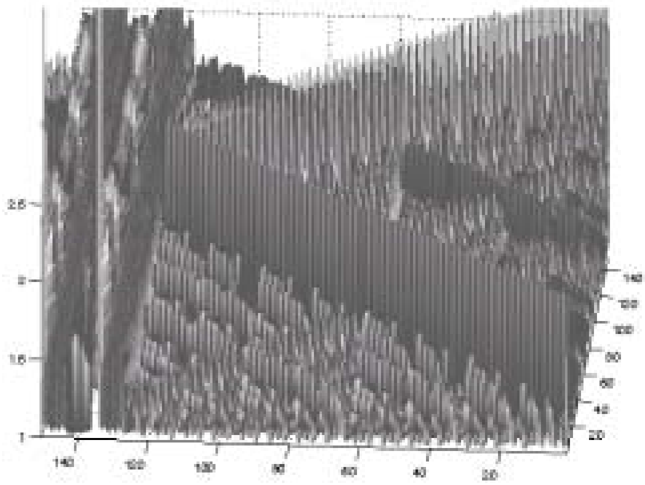

Figure 3. Landscape of Normalized Mutual Info.

Event	Fitness1	Fitness2	Fitness3
100% Funct.	13 ± 5	14 ± 7	18 ± 6
Opt. Soltn.	30 ± 7	30 ± 10	40 ± 20

Table 1. Results of Experiment 1

Event	Fitness1	Fitness2	Fitness3
100% Funct.	39 ± 12	40 ± 11	50 ± 12
Opt. Solutn.	160 ± 15	167 ± 15	170 ± 20

Table 2. Results of Experiment 2

6.1 Experiment 1

Here we design the following (simple) Boolean function:

$$F(a, b, c, d) = \sum(0, 1, 2, 3, 4, 6, 8, 9, 12) = 1$$

We use a population size of 300 individuals, $p_c = 0.35$, $p_m = 0.65$, and we run our algorithm for 100 generations. The optimal solution has 6 nodes, thus we find the generation in which the first 100% functional solution appears, and the generation number where the optimal is found. The problem was solved 20 times for each fitness function.

Table 1 shows the results of these experiments.

6.2 Experiment 2

The next test function is:

$$F(a, b, c, d, e, f) = ab + cd + ef$$

In this case, we use a population size of 600 individuals, $p_c = 0.35$, $p_m = 0.65$, and we stop after 200 generations. The optimal solutions has 14 nodes. Each problem was solved 20 times for each fitnesss function.

Table 2 shows the results of these experiments.

7 Final remarks and conclusions

A fitness function using only conditional entropy was tested with no success at all. We believe this is a clear indication of a fitness function that does not take into account the properties of entropy. In general, the three fitness functions work quite well, all of them find the optimum in most cases (with some higher probability than in previous experiments), thus comparable to other fitness functions based on Hamming distances. We introduced a combination of Hamming distance and information theory measures in the fitness function to guide the population towards only one attractor, avoiding the use of crowding methods in that way.

Based on the results shown in Tables 1 and 2 we would give some advantage to normalized mutual information over simple mutual information because it is less biased. The comparison of the fitness landscapes is encouraging.

Acknowledgements

The first author acknowledges partial support from CONACyT project No. I-39324-A. The second author acknowledges support from CONACyT through project 32999-A.

References

[1] Arturo Hernández Aguirre, Bill P. Buckles, and Carlos Coello Coello. Evolutionary synthesis of logic functions using multiplexers. In C. Dagli, A.L. Buczak, and et al., editors, *Proceedings of the 10th Conference Smart Engineering System Design*, pages 311–315, New York, 2000. ASME Press.

[2] Arturo Hernández Aguirre, Carlos Coello Coello, and Bill P. Buckles. A genetic programming approach to logic function synthesis by means of multiplexers. In Adrian Stoica, Didier Keymeulen, and Jason Lohn, editors, *Proceedings of the First NASA/DoD Workshop on Evolvable Hardware*, pages 46–53, Los Alamitos, California, 1991. IEEE Computer Society.

[3] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. John Wiley & Sons, New York, 1991.

[4] Van Hoyweghen, D. Goldberg, and B. Naudts. From twomax to the ising model: Easy and hard symmetrical problems. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO2002*, pages 626–633, San Francisco, CA, 2002. Morgan Kaufmann Publishers.

[5] C. Studholme, D.L.G. Hill, and D.J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32:71–86, 1999.