Computation in Genetic Code-Like Transformations

Hillol Kargupta
School of Electrical Engineering and Computer Science
Washington State University
Pullman, WA 99164-2752, USA
hillol@eecs.wsu.edu

The gene expression process in nature evaluates the fitness of a DNA through the production of different proteins in different cells. The DNA sequence first produces the mRNA sequence and the mRNA produces protein by using a transformation called the *genetic code*[1]. It has been shown [2]that genetic code-like transformations introduce very interesting properties to the representation of a genetic fitness function. Such transformations can convert functions[1] with exponentially large description in Fourier[2] basis to one that is highly suitable for polynomial-size approximation. Such transformations can construct a Fourier representation with only a polynomial number of terms that are exponentially more significant than the rest when fitter chromosomes are given more copies through a redundant, equivalent representation. This is a very desirable property for efficient induction of function representation from data which is a fundamental problem in learning, data mining, and optimization.

References

[1] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson. *Molecular biology of the cell.* Garland Publishing Inc., New York, 1994.

[2] H. Kargupta. A striking property of genetic code-like transformations. School of EECS Technical Report EECS–99-004, Washington State University, Pullman, WA, 1999. http://www.eecs.wsu.edu/~hillol/gene_expression.html.

[3] S. Kushilevitz and Y. Mansour. Learning decision trees using fourier spectrum. In *Proc. 23rd Annual ACM Symp. on Theory of Computing*, pages 455–464, 1991.

[1] Although the class of such functions is yet to be precisely characterized, the paper develops a general understanding.

[2] The analysis is identical to that using Walsh basis; however, I choose the term Fourier because of its historical [3] use in function approximation literature.