

Current Research at Home of Mathematical Genomics

Pabitra Pal Choudhury

Mathematical Genomics. Webpage: www.isical.ac.in/hmg

Tools employed: Integral Value Transformations, Fractal Geometry, Mathematical Morphology and Cellular Automata.

We have designed a class of transformations namely Integral Value Transformations (IVTs) which is essentially a generalization of special kind of Cellular Automata (CA). Like CA, we have induced Discrete Dynamical System (DDS) structure on \mathbb{N} through one dimensional IVTs.

In near future, we have aim to study about their dynamics of those DDS through Topological Dynamics and/or Complex Dynamics.

We have found out Collatz like function in the domain IVTs and already been explored the analogue results in our paradigm. I would like to unify the entire Collatz like functional space in the light of Dynamical behaviour. Most of the one dimensional IVTs are everywhere continuous, nowhere smooth in the discrete space \mathbb{N} . So, a special kind of sequences (self-similar fractal like sequences) can be generated through IVTs on which we can study for their quantitative characterization. The entire work can be done in k -dimensional space in light of Fractal and Dynamics. The discrete sequences have different application in Digital Signal Processing, Mathematics Genomics, and Cryptology and so on. We would like to explore possible scopes in the said application areas

We know fractal parameters namely fractal dimension, succolarity, lacunarity etc are responsible to give an insight of inner content like complexity (order/disorder) of a geometric object. To the best of our knowledge, there is no fractal parameter which is responsible to exhibit the morphometry of a geometric object (fractal). So it would not be surely possible to characterize a fractal, quantitatively using only fractal parameters because two different geometric objects can have same fractal content. It is our strong conviction that the morphological parameters (Bifurcation dimension, Morphological Entropy, Granulometric Analysis) along with fractal parameters can be used as a signature of fractal. In the near future we would like to establish the fact. It is essentially a geometric characterization of fractal in the light of Mathematical Morphology.

It is well known that DNA is the governing code for all organisms. DNA consists of four nucleotide bases A, T, C and G.

There are 64 triplets consisting of A, T, C and G. These are called *Codon*. All these triplets except TGA, TAA and TAG (to biologists as '*stop codons*') are responsible for protein formation. A DNA sequence consisting of stop codons is known as Intron (Pseudo gene i.e. these are not responsible for protein formation) and the rest are known as Exon (Functional gene). We (Human) are so fortunate that we have a class of genes (DNA) which are full of exons and these genes are the smallest in length too. The class is known as Human Olfactory Receptors (ORs). After Human Genome Project (by Govt of USA) it is deciphered that there are only ~ 700 total number of ORs (genes) of length on an average about 1000 bp (base-pair) in Human genome.

But it is true that we can make (4^{1000}) codes of length 1000 which consist of A, T, C and G. But out of these many staffs, only 700 are being selected (generated) by NATURE as human olfactory DNA sequences.

Now we are in front of some serious fundamental issues which are as follows:

NATURE might have used the olfactory receptors which are in our genome for olfaction purpose by either formation or selection or both. Now if it is by formation, then what is the formation methodology (s)? If it is by selection then what governs the selection process and what are the selection parameters? In other words, for selected Ors (which are available in Human genome) what are the specialties that they have which are not available in the rest?

If pseudo genes are not at all functional then why does NATURE carry these junks for prolong time over evolution? It should have been deleted over millions of years of biological evolution. We all agree that NATURE is very economic, SHE is the supreme economist. We have seen that there have been almost 50% pseudo genes available in the human olfaction family. Since they exist, we believe that they might be associated with the functional genes in some form or others (already such issued have been raised in research article). If so, what is their association and what are the association quantitative features coupled with it?

So we all agree that there are still many more A, T, C, G chains in the domain of (4^{1000}) which are almost like existing pseudo genes. But then, the question is what is so special in those pseudo genes about their non-functionality or pseudo characteristics? Are there other members having almost the same characteristics in the domain of (4^{1000})?

In recent past it reported that a genomic sequence (PTPN11/Shp2 of Human) is act as Oncogene (responsible for cancer) as well as it works as tumour-suppressor function in liver. This is a first kind of report published in the journal Cell. We have made an effort to characterize the genomic landscape in the gene sequence, quantitatively through Fractal and Mathematical morphology. In near future we would like to explore filtered the other such genes from various Genomes though our quantitative study.

The similar issued can be made for any other genome of different organisms (like Cancer Genome in Human). Till the date what we have presently is only few databases of DNA. We need to study them quantitatively for proper understanding of their function, evolution. We would like to study them through the looking glass of Fractal and Mathematical Morphology and other some mathematical and statistical tools. This study we definitely help in making a great success in Genomic Therapy/Genomic Medicine/ detection of genomic diseases.

For further details of Home of Mathematical Genomics and allied research works, please visit the link

<http://www.isical.ac.in/~hmg>

Note: This document is a verbatim copy of the research proposal written by Sk. Sarif Hassan.