MAP-I PhD Proposal

Title: Inter-Symbolic Distance Analysis in Genomic Signals

Keywords: Inter-Symbolic Distance, DNA, Genome, Motifs, Signal Processing, Information extraction

Supervisors:

Carlos Bastos, <u>cbastos@ua.pt</u>, University of Aveiro, Institute of Electronics and Telematics Engineering of Aveiro, IEETA/DETI

Vera Afreixo, <u>vera@ua.pt</u>, University of Aveiro, Center for Research and Development in Mathematics and Applications, CIDMA/DMAT

Research Unit: IEETA/UA

Introduction

This proposal is in the context of the work developed for some years by our research group in computational biology and the use of inter-symbolic distances to extract information from the genomic signals (sequences).

The inter-symbolic distances have already revealed potential in the identification of some functional/regulatory genome regions (e. g. coding regions [1] and CpG islands [2]). The distances have also been used to create genomic signatures and evolutionary phylogeny [3].

Objectives

The main aim of this PhD work is to develop analysis tools capable of extracting biological relevant information (for example: motifs, detection/prediction of patterns, regulatory elements, functional elements, structural elements and "universal" laws in DNA) from the genome using the inter-symbolic distances.

The tools should also fully characterize the inter-symbolic distance sequences for symbols of various lengths. The developed tools should be capable of characterizing the distances both globally (for the complete sequence) and locally (along the sequence).

References

[1] Carlos A C Bastos, Vera Afreixo, Sara Pinto Garcia, Armando J. Pinho. Inter-STOP symbol distances for the identification of coding regions. Journal of Integrative Bioinformatics, vol. 10, no. 3, p. 230, 2013.

[2] Hackenberg, M., Previti, C., Luque-Escamilla, P.L., Carpena, P., Martínez-Aroza, J., Oliver, J.L. CpGcluster: A distance-based algorithm for CpG-island detection. BMC Bioinformatics, 7:446, 2006.

[3] Vera Afreixo, Carlos A C Bastos, Armando J. Pinho, Sara Pinto Garcia, Paulo J S G Ferreira. Genome analysis with inter-nucleotide distances. Bioinformatics, vol. 25, no. 23, p. 3064-3070, 2009 (doi: 10.1093/bioinformatics/btp546).