

Biological Data Clustering

Supervisor: P. Fazendeiro

1 Abstract

The main purpose of this project is to fine tune a state of the art clustering algorithm in order to perform an “omics” data analysis task. The biological data analysis presents many challenges due essentially to the high-throughput and high-dimensional data. The work plan includes the set up of a data clustering experiment and selection of benchmark datasets, study of a particular clustering algorithm and its parametrization, the tuning and test of the clustering algorithm for the particular problem being tackled and, finally, the analysis of the results. The potential applications of the developed system depend entirely on the type of biological data chosen by the candidate.

2 Contact and Working Place

Enquires regarding this project should be addressed to Prof. Paulo Fazendeiro (fazendeiro@ubi.pt)

The experimental part of the work will be performed at SOCIA lab (<http://socialab.di.ubi.pt/>).

3 Objectives and Tasks

T1 To review the related literature

T2 To set up the experiment, select the training and test data

T3 To study a clustering algorithm and its parametrization

T4 To compare the performance of the different parametrizations

T5 To write up a MSc thesis and a scientific paper

4 Timetable

T1 1 month

T2 1.5 months

T3 1.5 months

T4 2 months

T5 2 months

5 Expected Results

- 1 journal paper
- 1 MSc thesis

6 References

P. Fazendeiro and J. Valente de Oliveira, Fuzzy Clustering as a Data-driven Development Environment for Information Granules, in Witold Pedrycz et al (Eds.), "Handbook of Granular Computing", Wiley, 2008.

P. Fazendeiro and J. Valente de Oliveira, A Fuzzy Clustering Algorithm with a Variable Focal Point, Proc. of IEEE World Congress on Computational Intelligence, WCCI 2008, Hong Kong, China, June, 2008.