Contribution ID : **3** Type : **not specified**

Parameters Estimation From Biomedical and Biochemical Data

- This poster abstract is intended to be submitted for the Dresden Summer School in Systems Biology apply -

I decided to give a generic overview of the main researches I worked on during my studies. The poster won't go too far into detail, trying to make it understandable for everyone.
It contains three works I pursued during my bachelor and master degree all related by the self-explaining title.

1) Dealing with biomedical signals like electrocardiograms (ECG) is not a trivial task. Entropy (and its related metrics) turned out to be a fundamental feature when we need to evaluate and classify possible anomalies of subjects. In literature several ways to estimate entropy from time series data can be found.
In this work, I'm investigating an alternative way to estimate this parameters by modeling the signal with an extension of the well known formalism Markov Chain called Variable Order Markov Model.

2) When a high level of detail is expected from quantitative predicitions of system dynamics, a fully parameterized, small-scale modeling approach needs to be adopted. Despite their highest predictive power, their applicability to complex biological systems is tipically limited by the lack of quantitative parameters. To overcome this problem, we chose to exploit an empirical biological knowledge (such as the phenotype of a cell) that helps to estimate the missing values.
The system is modeled by experts with an intuitive formalism called Stochastic Petri Net (SPN) which is then automatically translated into a set of Ordinary Differential Equations in which parameter uncertainty is overcome through an Optimization Problem whose objective function encodes a biological phenomena.
This approach has been exploited to study the metabolic behavior of cancer cells, overcoming the lack of experimental data, as normally happens in cancer cell scenarios.

3) DNA sequencing produces a huge amount of data which is hard to manage in data analysis. To overcome the well known curse of dimensionality problem, several algorithms for feature selection are available in literature.
In the genomics field, the task is harder, because it is important to ensure that the genes (features) selected are actually carriers of information, minimizing the probability of the data to be affected by sequencing methods, faulty data or a lack of statistical background.
In this work, by exploiting a statistical method called Negative Binomial Model, I developed a solid algorithm for feature selection among several samples coming from different populations.
This algorithm has been tested on a dataset of patients affected by one of four kinds of different cancers, whose inital feature dimension was over 20.000 genes. The algorithm properly reduced the dimension to 300 features, allowing any clustering/classification algorithm to run in a resonable amount of time and achieving an accuracy of classification close to 100%.

Summary

Primary author(s) : Dr AMELI, Corrado

Presenter(s) : Dr AMELI, Corrado

Session Classification : Poster Session