Abstract

This work addresses some of new technological frontiers in clinical biology that are at the intersection of logic, information theory, and statistical inference. The principal focus was the development of Bayesian Decision Support System engine based on the information theory and it was completed successfully. Author proves that his view is not only theoretical construct, but a practical reality. Some enablers and the vision for next generation clinical systems stands out clearly from this work. This thesis is also a valuable requirements repository that can serve well as the basis for commercial developments. Although the main trust was put into DNA microarray genotype to phenotype cross-correlations the solution presented goes beyond genotype to phenotype studies. The architecture developed addresses the critical clinical operational pressures from seamless integration to financial considerations like CAPEX/OPEX optimization. An original idea of Bayesian cancer signature -based on arch distributions - allowing prediction and diagnosis is promoted, herein. The idea is not restricted to cancer, but in this context is presented. Its main value result from the new perspective and analytical tools available for structure studies including some of Digital Signal Processing(DSP) methods. No other known literature studies network structure with the aid of the distribution functions and DSP. All of the tools implemented are used to analyze cancer and both structural and dynamic examples are given. In the text one finds new ideas and methodological questions being addressed clearly. For instance diagnostic generalization unifying logic and probability together with cancer synthetic construction method is developed. Certainly, the integration of engineering and biology is reality and this thesis shows how it is done. Some of the qualitative work has been accomplished in Princeton, New Jersey, USA and author thanks for the opportunity of doing so. It still remains to be seen whether the weighted information measures can provide value in genomic field but without any doubt the constrained learning based on distribution functions and implemented as part of this thesis provides new frontiers to uniform structure investigation and commonality identification especially important in the case of cancer.

Keywords: Cancer, DNA Micro-arrays, Proteomics, Statistics, Baye's rule, epistemology, diagnosis, prognosis, networks, machine learning, java, chaos, grammar, systems biology, logic