J Comput Sci Syst Biol 2016, 9:5 (Suppl)
http://dx.doi.org/10.4172/0974-7230.C1.015

Learning from uncertain kernel in bioinformatics

Sahely Bhadra
Aalto University, Finland

In this talk I will present the problem of designing machine learning methods when the biological data samples are contaminated with noise or there are missing values. Automatic kernelized classification techniques by studying structural or metabolic similarities among samples are popular for deriving a variety of insights, such as the ancestral links, the shared regulation or function of biochemical components. However, experimental measurements of biological processes and structures are subject to substantial noise stemming from both technical and biological variability. Again, different biological data sets provide information on different features and samples. Considering such multiple data sets for better learning poses the challenge that information for each point are not available in all data sets creating missing entries. First part of this talk is about the problem of designing SVM classifiers when the similarity measure or the kernel matrix, *K*, is affected by uncertainty, i.e., an expert is not fully certain about the value. In this work the uncertainty in the kernel matrix is modeled by a convex and bounded uncertain set which encompasses possible realizations of K. This new approach leads to an optimization problem which could be cast as a saddle point (minimax) problem. Due to favourable conditions satisfied by the problem, we are able to use a efficient algorithm which uses only the first order information, yet proved to decrease the initial error by a factor of $O(1/T^2)$, after T iterations. A comprehensive empirical study on real-world protein structure data sets show that the proposed formulations achieve the desired robustness to structural noise. The second part will discuss about a novel technique to complete missing rows and columns in the kernel matrix with the help of other available informations in case of multiple data sources. The proposed approach predicts missing (rows as well as columns) entries by modeling both within-view and between-view relationships among kernel values.

sahely.bhadra@aalto.fi

Refinement of predicted approximate 3-D structure of protein

Osmar Norberto De Souza
PUCRS, Brazil

Although the protein folding/structure problem is known since the early 1950s, the protein folding/structure prediction (PSP) problem was presented about 52 years ago and continues unsolved. PSP constitutes a major challenge in the structural biology and bioinformatics research fields. As a contribution to possible solutions to this problem, we have developed the CReF (Central Residue Fragment-Based) method for the prediction of the approximate 3-D structure of proteins. CReF can predict correctly the elements of secondary structure of proteins but not their native fold. For that reason, we employ molecular dynamics (MD) simulations with restraints on the main-chain torsion angle of regular secondary structures, and at different temperatures, to refine the CReF predictions. In this presentation, we will report improvements in our refinement protocols. Refinements with an RMSD of as low as 1.3 Å for a 53 amino acid residues protein can be attained in the 50 ns time scale, a far smaller molecular dynamics time scale than most current folding prediction protocols. Also, we will show results for small, medium and large proteins belonging to the α, β, and α/β structural classes.

osmar.norberto@pucrs.br