Poster I-80

XXStout: Improving the Prediction of Long range residue contacts

Authors:
Ian Walsh (*University College Dublin School of Computer Science and Informatics*)
Alessandro Vullo (*University College Dublin School of Computer Science and Informatics*)
Gianluca Pollastri (*University College Dublin School of Computer Science and Informatics*)

Short Abstract: This poster discusses our XXStout server, which is a state of the art residue contact predictor. It uses the novel feature of contact density as input, contact density is the principal eigenvector of the residue contact matrix. We show that using contact density significantly increases the performance.

Long Abstract:

XXStout: Improving the Prediction of Long range residue contacts
Ian Walsh, Alessandro Vullo, Gianluca Pollastri
School of Computer Science and Informatics, University College Dublin,
Belfield, Dublin 4, Ireland
{ian.walsh,alessandro.vullo,gianluca.pollastri}@ucd.ie

Motivation:
Accurate contact map prediction is a crucial intermediate step in predicting protein 3D structure. [1] have shown that it is possible to reconstruct the 3D conformation of a protein from accurate and even noisy contact maps. Also, because they are translational and rotational invariant, as opposed to the 3D structure, they are ideal for this intermediate stage. Prediction of long range residue contacts are the most difficult, unfortunately they are the most important since they determine the interactions of residues at large sequence separations (e.g. beta sheets).

Our 3D prediction server, Distill, uses contact information as one of its main constraints when building the protein backbone. Aside from 3D prediction, accurately predicted contact maps can be useful in many other potential applications such as database searches for a specific protein family and domain prediction.

Background:
In this poster we present our server XXStout which predicts the probability of residue i and j, of a sequence of length N, having a distance in 3D between [0,T], where T is the cutoff threshold in Angstrom. The server allows us to choose T at 6, 8 and 12 Angstrom. It will return a NxN probability matrix, where the contact map is defined as the binary matrix, where we choose (i,j) to be in contact at a give probability threshold, otherwise (i,j) is considered non-contact.

Map predictions are got from using using multiple alignments, secondary structure (via our Porter server [2]), relative solvent accessibility (via our PaleAle server [3]) and the predicted contact density (via BrownAle [4]). Particular attention is paid to the role of the contact density predictor and we conclude that the addition of this input feature increases the performance of XXStout, especially for long range contacts.

BrownAle essentially predicts the Principal Eigenvector (PE) of the contact map. [5] have developed a branch and bound algorithm that can reconstruct globular proteins which are <= 120 amino acids in length from accurate values of the PE components. This reconstruction is surprising considering there are possibly N-1 other eigenvectors of the contact map. It indicates that perhaps the eigenvector is a good 1D encoding of the 2D contact map and hence its 3D topology. The PE is a sequence of length N, where N is the length of the amino acid sequence and is related to contact density of a particular residue. Algorithms using 1D-RNN's [2][3][4][6] have being used in similar tasks for predicting sequences of the same length as the amino acid sequence.

The inputs presented to the BrownAle server are amino acid alignment profile, secondary structure indicator (helix, coil, strand) and relative solvent accessibility at residue i in the sequence, the length of the protein is also presented. The predictions made by BrownAle are then used as an input feature in XXStout. We use 4 classes for this input, where 0 indicates low contact density and 3 indicates a high contact density.

Results:
BrownAle's predictions are significantly above a baseline predictor, while XXStout's CASP style measures of precision (accuracy) recall (coverage) and F1 show improvements, especially for the crucial long range contacts [4].

XXstout's position in the overall pipeline of our main ab initio protein structure predictor, Distill, is presented. We demonstrate that BrownAle's predictions can clearly mimic actual contact density. Also, some predictions are presented showing the clear difference between maps using BrownAle and Maps not using BrownAle.

Conclusion:
Our final predictor for contact maps achieves state of the art performance when using BrownAle (contact density predictor) as input along with secondary structure, relative accessibility and multiple alignments. A suite of predictors of structural features, including the contact density, and contact density based contact maps, is available at http://distill.ucd.ie.

Refereneces:
[1] Porto M, Bastolla U, Roman H, Vendruscolo M: Reconstruction of protein structures from a vectorial representation. Phys.Rev.Lett. 2004, 92:218101.

[2] Pollastri G, McLysaght A: Porter: a new, accurate server for protein secondary structure prediction. Bioinformatics 2005, 21(8):1719–20.

[3] Pollastri G, Fariselli P, Casadio R, Baldi P: Prediction of Coordination Number and

Relative Solvent Accessibility in Proteins. Proteins 2002, 47:142–235.

[4] A. Vullo, I. Walsh, G. Pollastri*. "A two-stage approach for improved prediction of residue contact maps" BMC Bioinformatics, 7:180, 2006.
Open access abstract and PDF (BMC Bioinformatics web site).
[5] Vendruscolo M, Kussell E, Domany E: Recovery of protein structure from contact maps. Folding and Design 1997, 2:295–306.

[6] Baldi P, Pollastri G: The Principled Design of Large-Scale Recursive Neural Network Architectures – DAG-RNNs and the Protein Structure Prediction Problem. Journal of Machine Learning Research 2003, 4(Sep):575–602.