べき乗則モデリングとサンプリングによるタンパク質複合体の予測手法の研究

丸山　修*

多くのタンパク質の機能は，特定のタンパク質複合体を通して発揮される．ゆえに，タンパク質複合体の同定は生命科学研究におけるインフラ的な基礎知識として必要不可欠である．それ故，情報科学的にタンパク質複合体を予測する手法が盛んに研究されている．

我々は，タンパク質複合体予測問題に対して，タンパク質間相互作用の重みに基づく最適化項に，二つの正則化（罰則）項を加えることにより評価関数を定式化し，これをマルコフ連鎖モンテカルロ (MCMC) 法の一種であるメトロポリス・ヘイスティングス法に基づくサンプリング・アルゴリズムで最適化する手法を設計した（[1, 2] ）．

このアルゴリズムが出力する解は，入力として与えられるタンパク質間相互作用データの全タンパク質からなる集合の分割である．この分割のサイズ 2 以上の個々の部分集合が予測されたタンパク質複合体である．

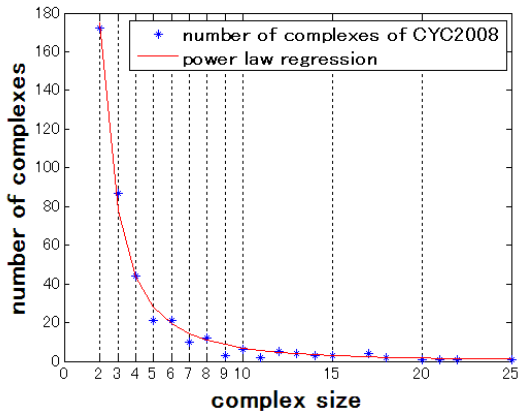

図　1:　酵母のタンパク質複合体データベース CYC2008 の複合体サイズ分布．

１つ目の正則化項は，タンパク質複合体の構成タンパク質の個数は「べき乗則」に従うというヒトや酵母のタンパク質複合体のデータベース解析の結果に基づくものである（図 1）．さらに，予測精度向上のため，全ての予測タンパク質複合体に含まれるタンパク質の総数を制御する正則化項を正規分布でモデリングした．これが第２の正規化項である．

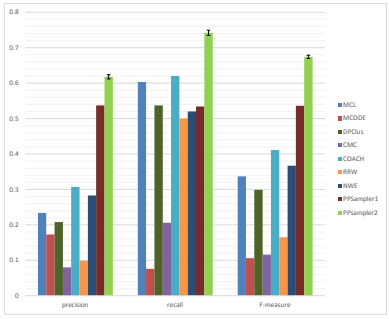

図 2: 精度比較.

この評価関数を最適化するサンプリング・アルゴリズムを実装した予測ツール PPSampler (Proteins' Partition Sampler) と既存手法との比較実験の結果，酵母に関しては，PPSampler は既存手法に対して 30 %以上の予測精度の向上を実現し（F 値 0.54）[1]．さらに評価関数と最適化アルゴリズムの改良を行い，F 値を 0.54 から 0.67 へと 27%向上させている（ツール名 PPSampler2 [2]）．

参考文献

[1] Daisuke Tatsuke and Osamu Maruyama. Sampling strategy for protein complex prediction using cluster size frequency. *Gene*, 518:152–158, 2013.

[2] Chasanah Kusumastuti Widita and Osamu Maruyama. Ppsampler2: Predicting protein complexes more accurately and efficiently by sampling. *BMC Systems Biology*, 2013, To appear.

*九州大学マス・フォア・インダストリ研究所