

Maximal Information Coefficient-based Oscillation Prediction による周期と振幅の定量

BI

井内仁志

背景

概日リズムは菌類から哺乳類まで幅広い生物種に内在するシステムである。近年の研究から、幾つかの精神疾患や睡眠障害が概日リズムに関係していることが知られている。ゆえに、トランスクリプトームデータ、プロテオームデータやメタボロームデータのような巨大なデータから振動物質を抽出することは、生命システムの理解や医療分野において非常に重要である。しかしながら既存の手法は、サンプリングポイントが比較的少ない生物学的データにおいて適用することが難しかったり、疑陽性が高い、位相を変えた多数のサインカーブと比較する必要があり、計算コストが高いなどの問題点があった。

結果

そこで我々は 2 変数の相関が非線形でも検出できる Maximal information-based nonparametric exploration 法 (MINE 法) を用いて、新規振動物質計算法を開発した。本手法はピアソン相関係数を用いた cross-correlation 法に比べて、位相を変えた多数のコサインカーブを準備する必要がないため処理が簡便であり、かつ疑陽性率が低いことを示した。さらに、Maximal information coefficient (MIC) を計算する際に必要な α の値やノイズが検出力に与える影響を検証した。最後に本手法を既に発表されている時系列プロテオームデータに適用したところ、新たな統計的有意に振動している物質を抽出することができた。

まとめ

本論文では疑陽性率が低く、かつ計算コストを抑えた振動物質検出法を開発した。また、ノイズや MIC を計算する際に必要な係数を変化させた時に結果にどのような影響を与えるかを検証した。その手法を公開されている時系列プロテオームデータに適用したところ、新規振動物質候補を検出できた。これらのことから、今回我々が開発した手法は有用であると考えられる。

なお、この成果は論文投稿中であるため、詳細は省略します。