

細胞の状態遷移検出のための 小サンプル時系列データの最適モデリング

富 永 大 介^{†1}

遺伝子発現の時系列を複数のステージに分け、分布モデルを各ステージに当てはめると、時系列全体の尤度を計算できる。また尤度が最大になるように分割点を選ぶことができ、何分割すればよいかを考慮に加えれば、情報量規準により最適な分割数を定めることができる。DNA microarray の時系列データの場合には、各観測時刻についてそこが分割点となる遺伝子の個数を、細胞レベルから見た状態遷移としての度合いとして見る事ができる。このアルゴリズムを実装し、シミュレーション・データと遺伝子発現データに適用した。

Optimal modeling method for small sample time series data for detection of state transition of cells

DAISUKE TOMINAGA^{†1}

Optimal division points on a gene expression time series data is defined by division points and optimal mathematical models of data distribution on each divided time series. The optimal number of division is also determined by the Information Criterion. Considering DNA microarray data, the number of genes that its time series data is divided on a point implies magnitude of a change of cell state on the point. We implemented the algorithm and analyzed simulated time series data and published gene expression time series dataset.

1. はじめに

時系列データ中で遷移が生じているかどうかを検出する方法はいくつかある。しかしたと

えば ANOVA を使った方法¹⁾ ではサンプルの時間的順序を考慮していない。ガウス過程の混合モデルを使った方法²⁾ も同じで、また混合されるモデルの寄与が明確に違わない場合には、解析が困難になる。単純にデータを分割してそれぞれ分布モデルを当てはめ、平均値の違いを検定する方法もあるが、これは分割をしない場合と直接比較して論ずることはできない。

そこで我々は、正規分布モデルと情報量規準を使って最適ステージ分割の有無とその時刻を求めるアルゴリズムを考案した。我々はこれまでに、サンプル数が少ない場合に、情報量規準を使った時系列データのモデル選択が有効であることを示している³⁾。つまり、時系列データを前部、後部に分割し、それぞれに分布モデルを当てはめ、データ全体の尤度を計算する。尤度が最大となるのはどの時点でデータを分けたときかを探索し、そこを遷移時刻であるとする(正確には、隣接する二つの観測時刻には含まれた時間内と判断される)。そして分割数を自由度として情報量基準を計算する。今回我々は、時系列データを2分割と分割しない場合を同時に考慮して最適分割を求めるプログラムを作成した。

2. 結果と結論

正規分布乱数によるシミュレーション・データでは、正規分布の分散はすべて1、データ長は12点、最短ステージ長は3とした。時系列データは500セット用意し、分割の場所は最短ステージ長より短くならない箇所にした。その結果、期待値が分散の2倍程度変化すれば、それを60%の割合で検出できることを確認した。

次に線虫の発生期のデータ(16分割までの10点のデータ)⁴⁾を解析し、2細胞期~4細胞期までは変動が少なく、最初の卵割および4細胞期から8細胞期に移行するとき大きな状態変化が生じていることが明らかになった。既報の論文⁵⁾では、4細胞期まではトランスクリプトームはあまり変化せず、これ以降原腸陥入までの変化が激しいとしており、4細胞期以降に関してはこれと一致する解析結果を得た。最初の卵割まではサンプルが少ないこともあり、解釈にはより慎重な検討が必要である。

参 考 文 献

- 1) Conesa A., et. al., *Bioinformatics*, **22**(9), pp. 1096-1102 (2006).
- 2) Stegle O., et. al., *Journal of Computational Biology*, **17**(3), pp. 355-367 (2010).
- 3) Tominaga D., *Biology and Bioinformatics Insight* (2010).
- 4) Baugh LR., et. al., *Development*, **132**, pp. 1843-1854 (2005).
- 5) Baugh LR., et. al., *Development*, **130**, pp. 889-900 (2003).

^{†1} 独立行政法人産業技術総合研究所, 生命情報工学研究センター
Computational Biology Research Center, AIST