

非線型変数領域選択手法の開発およびソフトセンサーへの応用

○金子 弘昌、船津 公人

東京大学大学院工学系研究科(〒113-8656 東京都文京区本郷 7-3-1)

【緒言】

スペクトル解析やプロセス管理などの分野における回帰分析において、隣接する説明変数 \mathbf{X} の間の相関が高いデータを扱う場合に変数を領域単位で選択する試みがなされている。変数領域選択手法の一つである genetic algorithm-based wavelength selection (GAWLS) 法 [1] は、GA を用いてモデルの予測性を表わす q^2 値を大きくする \mathbf{X} の変数の組を領域単位で選び出す変数選択法であり、変数領域とその組み合わせを同時に最適化することが可能である。しかし GAWLS 法におけるモデル構築手法として線型回帰分析手法の一つである partial least squares (PLS) 法が使用されているため、 \mathbf{X} と目的変数 \mathbf{y} の間に非線型関係がある場合において対応が不可能となる。そこで我々は、変数間に非線型性が存在する場合においても適切な変数領域選択と予測的なモデル構築を同時に達成することを目的として、GAWLS 法と非線型回帰分析手法の一つである support vector regression (SVR) [2] を組み合わせた変数領域選択手法(GAWLS-SVR 法)を開発した[3]。

【手法】

GAWLS-SVR 法とは、GAWLS 法における PLS を SVR とした手法である。染色体の適合度として、染色体に表現された波長領域について SVR 解析を行った後の q^2 値を用いることで、予測性に優れた非線型モデルを構築できる変数領域の組が得られるといえる。しかし SVR 法においては最適化すべきパラメータが3つも存在するため、GAWLS における PLS 法と同様にクロスバリデーションによりパラメータを最適化した後に q^2 値を求めると多大な時間がかかってしまう。そこで SVR モデルのパラメータも GA の染色体に含めることにより計算時間の問題に対応する。図1に GAWLS-SVR 法の概念図を示す。本手法では染色体の前半で変数領域を表現し、最後の3つの成分で SVR パラメータを表現する。これにより SVR モデルの q^2 値を計算する際に3つのパラメータを最適化する必要がなく、計算時間を大幅に短縮可能である。本手法を用いることで、 \mathbf{X} と \mathbf{y} の変数間に非線型性が内在する場合においても変数領域の組と SVR モデルを同時に最適化できる。

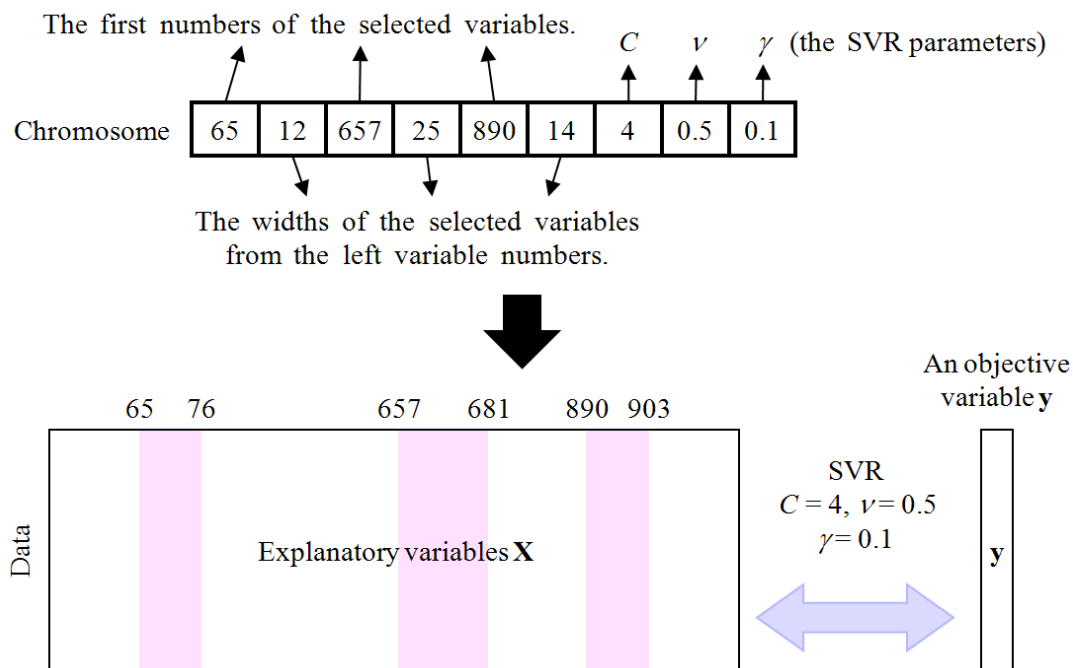


図1. GAWLS-SVR 法の概念図 (領域の数が3つの場合)

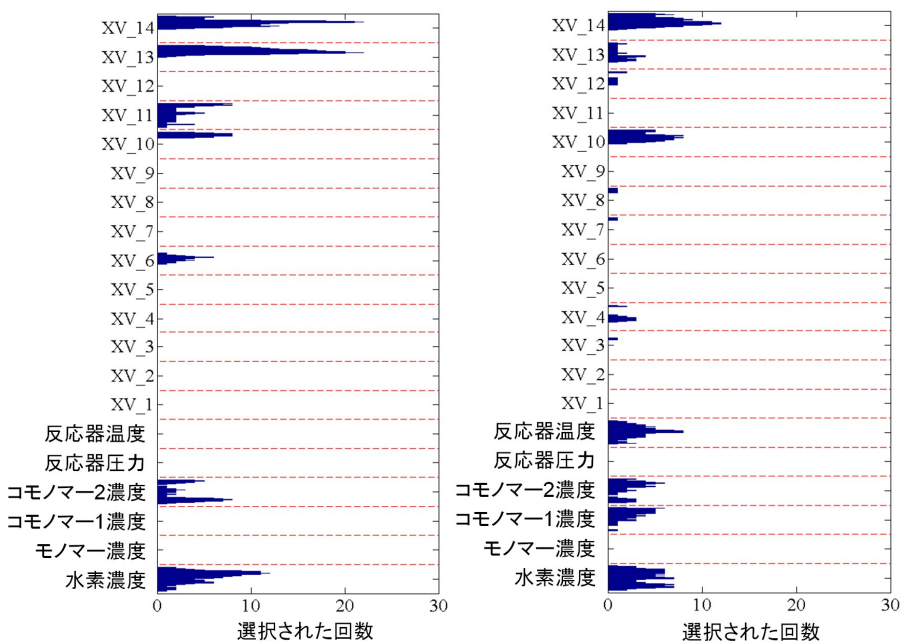
【結果と考察】

シミュレーションデータを用いた解析を行い提案手法の有用性を確認した後、本手法をソフトセンサーへ応用した。ソフトセンサーとは、オンラインで測定可能な変数と測定困難な変数の間で数値モデルを構築し、目的とした変数の値を推定する方法である。今回は三井化学市原工場のポリマー重合プラントで実際に測定されたデータを用いた解析を行った。対象としたプロセスにおいては、ポリマー物性とその他のプロセス変数間の非線型性が原因の一つとなり、予測的なソフトセンサーモデルの構築は困難となっている。そこで我々は反応器温度およびモノマー・コモノマー濃度等の 20 のプロセス変数から melt flow rate (MFR) を推定するソフトセンサーモデルを構築した。今回はプロセスの動特性を考慮するため、各プロセス変数がある時間まで遅らせた変数を \mathbf{X} に追加して最終的な \mathbf{X} とした。

それぞれの手法により 30 回モデル構築を行った結果、GAWLS モデルの r^2 と q^2 の平均値がそれぞれ 0.968 と 0.967 であり、GAWLS-SVR モデルのそれらの値がそれぞれ 0.995 と 0.992 であった。提案手法を用いることで、GAWLS と比較して精

度と予測性の高いモデルが構築されたことが分かる。SVR 法により \mathbf{X} と \mathbf{y} の間の非線型性に対応できたためと考えられる。

GAWLS 法 および GAWLS-SVR 法を用いた際の各変数の選択された回数を図 2 に示す。GAWLS 法では選択されなかった反応器温度やコモノマー濃度に対応する領域が GAWLS-SVR 法により多く選択された。これらの変数は MFR と関連の深い変数であることから、提案手法を用いることで適切な変数選択が可能であることを確認した。



(a) GAWLS 法

(a) GAWLS-SVR 法

図 2. 変数ごとの選択された回数

【結言】

本研究では、 \mathbf{X} と \mathbf{y} の間に非線型性が存在する場合においても適切な変数選択とモデリングを行うことを目的として、GAWLS 法と SVR 法を組み合わせた新規な変数領域選択手法を提案した。本手法をソフトセンサーに応用した結果、 \mathbf{X} と \mathbf{y} の間の非線型性を考慮した適切な変数選択が達成され精度と予測性能の高いモデルが構築可能であることを確認した。本手法により少数の \mathbf{X} の変数のみでモデリングが行われることで、高い予測性能と解釈のしやすさを併せ持つ実用的なモデルを得ることができると考えられる。

【謝辞】

データ提供の面で多大なご支援をいただいた、三井化学株式会社に深く感謝致します。

【参考文献】

- [1] Arakawa M, Yamashita Y, Funatsu K. *J. Chemom.* 2011;25:10–19.
- [2] Vapnik V.N. *The Nature of Statistical Learning Theory, second ed.*, Springer, New York (1999).
- [3] 金子 弘昌, 船津 公人, *J. Comput. Chem., Jpn.* 2011;10:122–130.