

ビッグデータから製品品質を精度良く予測できる NCLMを用いたソフトセンサー技術

Nearest Correlation Louvain Method (NCLM) for Soft Sensor Modeling Capable of
Accurately Estimating Product Quality from Big Data

業界共通の課題であるソフトセンサーモデルの 予測精度向上を実現

ビッグデータ化やIoT (Internet of Things) 化の進展で、業界では大量のデータの活用が進み、特に、測定が困難な製品の品質については、測定が可能な変数から予測するソフトセンサー技術が広く使われています⁽¹⁾。しかし、予測モデルに用いる変数を、大量の変数の中から見付けることが困難といった課題があり、また、更なる品質の改善には予測精度の高いモデルが望まれています。そこで東芝は、変数選択にNCLM (Nearest Correlation Louvain Method) を適用するソフトセンサーモデルを構築しました。加工設備の異常検知や工程管理への活用では、NCLMによる変数選択をした場合、従来の変数選択をしなかった場合に比べ、予測精度が約11～44%、誤差が約4～7%改善され、このソフトセンサーモデルの有効性を確認しました。

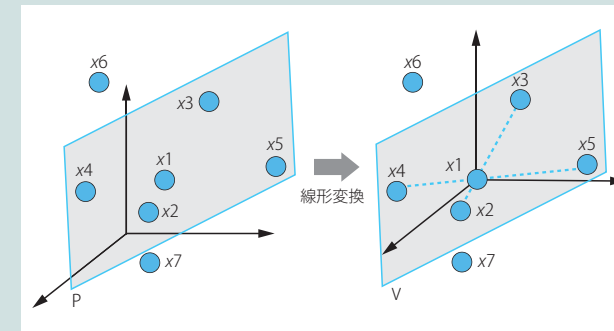
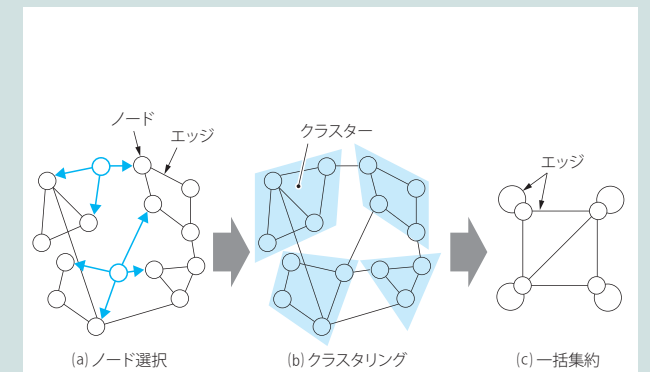


図1. NCの概要
空間Pから空間Vへ線形変換することで、大量のデータの中から類似した変数を見付け出せます。

表1. 評価結果

評価手法	基本手法	改善度(%)	
		予測精度	誤差
NCLM+PLS	SW+MR	44.46	7.16
	PLS	35.31	5.15
NCLM+LW-PLS	LW-PLS	32.42	5.37
NCLM+SVR	SVR	10.88	4.14

SW + MR: Stepwise+Multiple Regression



$$Q = \sum_{i \in C} \left\{ \frac{e_{ii}}{2m} - \left(\frac{\sum_{j \in C} e_{ij}}{2m} \right)^2 \right\}$$

Q: モジュラリティー
m: グラフ中の総エッジ数
C: クラスター集合
e_{ij}: クラスターiとjの間のエッジ数

図2. LMの概要

モジュラリティーQが最大になるまでクラスタリングすることで、目的変数を、最も精度良く説明できる組み合わせに分割できます。

ソフトセンサー技術の産業展開とその課題

ビッグデータ化が進む中、社会インフラや、エネルギー、ストレージ、製薬、化学、石油といった分野において、ソフトセンサー技術を加工設備の異常検知や、工程管理、設備特性の予測などに活用できないか、技術開発や検討が進められています。ある製品への適用を考えると、製造プロセス内の出来栄や加工設備情報を説明変数とし、それらの変数を使って最終的な製品の品質に影響する品質特性値(目的変数)を予測することになります。そして、その品質特性値が管理値内に収まるように説明変数を制御することで、製品の品質管理が行えます。しかし、予測モデルに用いる変数を、大量の説明変数の中から見付けることは困難といった課題があり、また、一般的に広く利用されている重回帰分析やPLS (Partial Least Squares Regression) などの予測モデル構築手法では、予測精度が低く、更なる品質の改善には予測精度の高いモデルを用いた品質特性値の管理が望まれています。

モデルの予測精度向上

東芝は、変数選択を行う手法の一つとして知られているNCLM^{(2), (3)}を用いて変数を選択し、予測精度が高いモデルを

構築する手法を開発しました。NCLMは、NC (Nearest Correlation)とLM (Louvain Method)の二つで構成されています。

NCは、着目する変数と同じ相関関係を持つ変数を見付け出す手法であり、図1に示すように、空間Pを変数x1が原点となるような空間Vへ線形変換し、各変数と原点を結ぶ直線上に、そのほかの変数が存在するかどうかを確認します。図1の例では、変数x2とx3、変数x4とx5がこれに相当します。これらのペアの相関係数は±1の値を取りますが、空間V上には変数x6やx7には、相関係数が±1となるようなペアは存在しません。相関係数が±1の変数ペアは変数x1と同じ相関を持つと判断しますが、実際には相関係数が厳密に±1となるような変数ペアを見付けることは困難です。このため、相関係数のしきい値をγ (-1 < γ < 1)として、相関係数がγを超えた場合、変数x1と同じ、あるいは類似した相関を持つとします。これを変数の数だけ繰り返すことによって、原点とする変数と類似した変数を探索することができます。

LM⁽⁴⁾は、グラフィカルクラスタリング手法の一つです。図2に、LMにおけるクラスタリングの手順を示します。まず、ランダムな順序でノードを選択し(図2(a))、モジュラリティーQが最も大きくなるノードを同一クラスターとします(図2(b))。次に、同一クラスター内のノードとエッジを一括集約して重み付

きグラフに変換します(図2(c))。この手順をモジュラリティーQが増加する限り繰り返します。最終的に、モジュラリティーQが最大値となったときのクラスタリング結果を採用します。

有効性の評価

今回は、NCLMの有効性を評価するにあたって、以下のサンプルデータを用意しました。

- 目的変数 : 製造工程内の中間出来栄(1個)
- 説明変数 : 製造工程内で使用した設備データ(70個)
- サンプル数 : モデル構築用(270個)、評価用(250個)

評価方法としては、NCLMを用いて変数の選択を行い、選択した変数を使って線形予測モデルと非線形予測モデルのそれぞれを構築し、NCLM非適用のときと比較しました。具体的には、線形予測モデルでは重回帰分析やPLSを、非線形予測モデルではSVR (Support Vector Regression) やLW-PLS (Locally Weighted PLS) を用いました。この結果、表1に示すように、NCLMによる変数選択をした方が、変数選択をしなかった場合と比較して、予測精度で約11～44%、誤差で約4～7%改善されることを確認しました。

今後の展望

ソフトセンサーモデルの予測精度を向上させる手法として、NCLMによる変数選択の導入が有効であることが分かりました。NCLMは、PLSなど対象となる予測モデル構築手法を選ばないので、広い範囲での応用が期待できます。今後、ソフトセンサーとして幅広い産業への展開を進めていきます。

文献

- Kano, M.; Fujiwara, K. Virtual Sensing Technology in Process Industries: Trends and Challenges Revealed by Recent Industrial Applications. J. Chem. Eng. Jpn. 2013, 46, 1, p.1-17.
- Uchimaru, T. et al. Nearest Correlation Louvain Method for Fast and Good Selection of Input Variables of Statistical Model. IFAC-PapersOnLine. 2015, 48, 8, p.123-128. <http://www.sciencedirect.com/science/article/pii/S2405896315010356>. (accessed 2017-11-30).
- 水町悠貴, ほか. "Nearest Correlation Louvain Methodを活用した予測モデル構築". The 4th Multi-symposium on Control Systems MSCS 2017. 岡山, 2017-03. 計測自動制御学会, 2017, 1B3-1. (USBメモリー).
- Blondel, V. D. et al. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment. 2008, 2008, 10, P10008.

曾我 朗

研究開発本部
生産技術センター 設計生産システム変革推進部