

教育研修・中規模データ分析のための機械学習ソフトウェア

「JUSE-StatWorks/V5 機械学習編」の機能と使い方

(株) 日本科学技術研修所
数理事業部 犬伏 秀生

本発表では、弊社で現在開発中の機械学習ソフトウェア「JUSE-StatWorks/V5 機械学習編」の機能と使い方をご紹介します。

※本内容は2018年11月時点の内容となります。今後予告なく変更する場合がありますので、ご承知おき下さい。

1. システムの概要

(1) 開発の背景

近年、機械学習が非常に高い予測性能を持つことが広く知られるようになり、様々な分野において機械学習の活用が進められています。品質管理分野でも、IoT（Internet of Things：モノのインターネット）などにより製造工程や市場から大量のデータを収集できる仕組みが整えられつつある中で、そのような大量データを利活用するための手段として機械学習が注目されています。

このような状況の下、品質管理分野における教育や実務で使用できる機械学習ソフトウェアの需要も高まっており、この度弊社では、JUSE-StatWorks/V5 シリーズの一つとして「JUSE-StatWorks/V5 機械学習編」（以下、StatWorks/V5 機械学習編）を新規にご提供する運びとなりました。

(2) システム概要

StatWorks/V5 機械学習編は品質管理分野をメインターゲットとしていますが、日本の品質管理分野の特徴の一つとして、問題解決や課題達成への取り組みの中で現場の技術者が自らデータ分析を行う点が挙げられます。このため品質管理分野では、今後、機械学習に関しても、専門部門だけではなく、各部門の現場の技術者が活用することが求められるようになることが予想されます。そして、各部門の現場の技術者が機械学習を活用できるようになるためには、多変量解析手法や実験計画法などの従来の SQC 手法と同様に、機械学習も各社の教育体系に組み込まれる必要があります。

以上の状況を踏まえ、StatWorks/V5 機械学習編は、StatWorks ユーザーを中心に、研修や現場の技術者が実務で使用することを想定し、次の特徴を持つシステムとしました。

【StatWorks/V5 機械学習編のシステム上の特徴】

- ① 教育や実務での使用の利便性のために、クラウドやサーバー上ではなく、クライアント PC 上で動作します。
- ② 全ての分析はマウス操作で行うことができます。スクリプトの記述は不要です。
- ③ StatWorks/V5 シリーズの既存製品と一体で使用できます（機械学習編のメニューは既存製品（総合編など）のメニューと統合されます）。
- ④ データの読み込み・編集は現行の StatWorks/V5 のワークシート上でを行い、機械学習の分析は StatWorks/V5 本体とは独立した機械学習モジュール上で行われます。
- ⑤ 機械学習モジュールは、既存の StatWorks/V5 と同様の画面・機能構成としており、StatWorks/V5 ユーザーは違和感なく使用できます。

(3) 搭載手法

StatWorks/V5 機械学習編には、次の機械学習手法が搭載されます。

| No | | 分類 | 解析手法 |
|----|-----------------|-----------|------------------|
| 1 | StatWorks 本体 | データクリーニング | データクリーニング |
| 2 | 機械学習編 モジュール | データ可視化 | 濃淡散布図 |
| 3 | | | 密度プロット |
| 4 | | | 等高線図 |
| 5 | | 情報要約 | カーネル主成分分析 |
| 6 | | | 混合ガウス分布 |
| 7 | | 正則化回帰 | リッジ回帰 |
| 8 | | | lasso 回帰 |
| 9 | | | Elastic Net |
| 10 | | 分類・予測 | サポートベクターマシン(SVM) |
| 11 | | | ランダム・フォレスト |
| 12 | 外れ値検出 | 1 クラス SVM | |
| 13 | 因果分析 | glasso | |



(4) 機能の特徴

StatWorks/V5 機械学習編の機能の特徴は次の通りです。

| No | 項目 | StatWorks/V5 機械学習編の特徴 |
|----|-------------|---|
| 1 | 分析可能なデータサイズ | StatWorks/V5 機械学習編では、分析対象データの入出力は既存の StatWorks/V5 本体のワークシート上で行いますので、分析可能なデータサイズはなります。特に、いわゆるビッグデータをそのまま StatWorks/V5 機械学習編で分析することはできませんが、いわゆるビッグデータから抽出したデータを用いてモデルの検討・評価を行うためには StatWorks/V5 機械学習編は有効であると認識しております。 |
| 2 | 分析結果の出力 | StatWorks/V5 機械学習編の分析結果は、既存の StatWorks/V5 の分析結果と同様に、グループとタブに整理して出力されます。このとき、グループの並びはおおよそ分析手順になります。 また、データや分析結果はできるだけグラフ化し、視覚的に確認できるようにしています。 |
| 3 | モデル選択 | StatWorks/V5 機械学習編に搭載した機械学習手法にはモデルの性能に影響を与えるハイパーパラメータが存在し、分析時にはそのハイパーパラメータをどの値にするのかを決める必要があります (StatWorks/V5 機械学習編では、ハイパーパラメータのチューニングを“モデル選択”と称します)。 StatWorks/V5 機械学習編では、このモデル選択を前面に出し、ハイパーパラメータがどのように決められているのかをユーザーが把握できるようにしています。 |

2. 搭載手法の概要

StatWorks/V5 機械学習編に搭載される機械学習手法の概要は次の通りです^{[1][2][3][4]}。

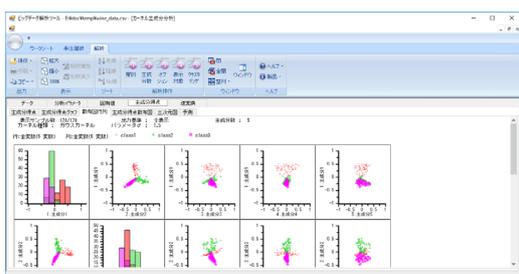
(1) データクリーニング
 平均値等による欠測の補完, マハラノビス汎距離による外れ値の除去などを行うことができます



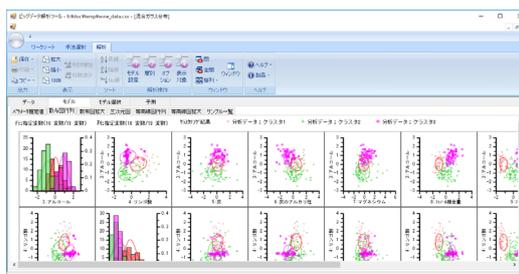
(2) 濃淡散布図・密度プロット・等高線図
 サンプル数が多い場合でもデータの分布状況を視覚的に確認できます。



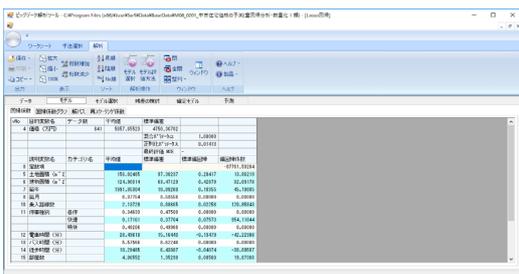
(3) カーネル主成分分析
 サンプルの分類を通常の主成分分析に比べて適切に行うことができます。



(4) 混合ガウス分布
 データの分布を複数の正規分布の重ね合わせで表すことができ、それに基づいて層別できます



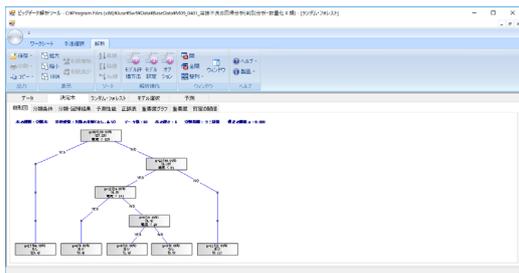
(5) リッジ回帰・lasso 回帰・Elastic Net
 説明変数間に多重共線性が発生している状況でも回帰式の当てはめや変数選択ができます。



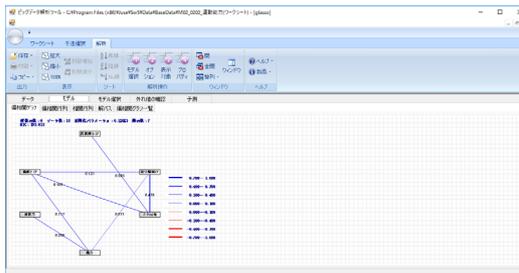
(6) サポートベクターマシン(SVM)・1クラス SVM
 目的変数(量的・質的)をより良い精度で予測することができます。



(7) ランダムフォレスト
 複数の決定木から、目的変数(量的・質的)の予測をより良い精度で行うことができます。



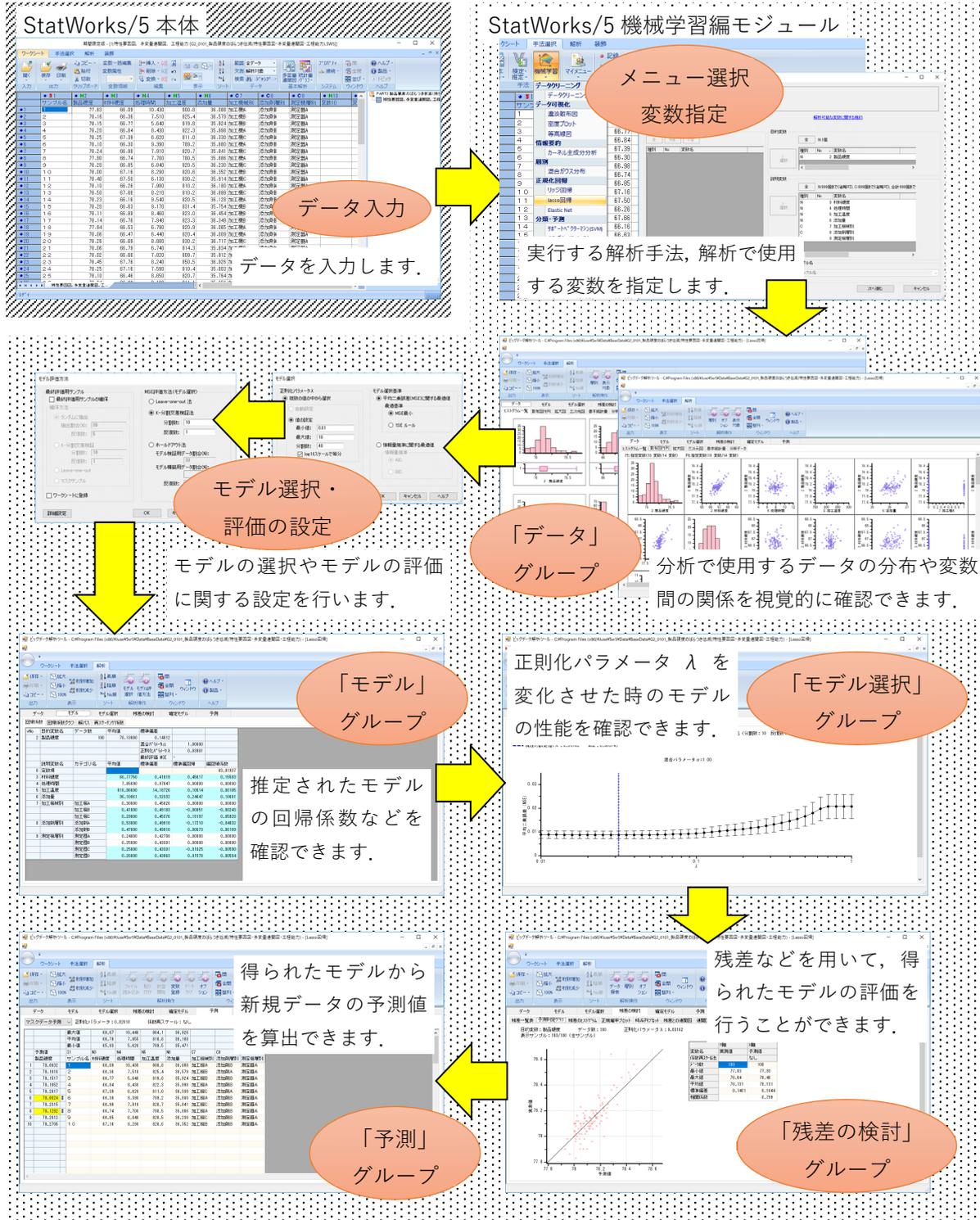
(8) glasso
 量的変数間の構造を自動的に構築し、それを視覚的に確認することができます。



3. 解析手法の機能構成例

ここでは、StatWorks/V5 機械学習編の解析手法「lasso 回帰」と「ランダム・フォレスト」の機能構成をご紹介します。

(1) 解析手法「lasso 回帰」



(2) 解析手法「ランダム・フォレスト」

StatWorks/5 本体

データ入力
データを入力します。

StatWorks/5 機械学習編モジュール

メニュー選択
変数指定

実行する解析手法, 解析で使用する変数を指定します。

モデル選択・
評価の設定

モデルの選択やモデルの評価に関する設定を行います。

「データ」
グループ

分析で使用するデータの分布や変数間の関係を視覚的に確認できます。

「決定木」
グループ

樹形図や正誤表など決定木の分析結果を確認できます

「ランダムフォレスト」
グループ

正誤表や重要度などランダムフォレストの分析結果を確認できます

「予測」
グループ

得られたモデルから新規データの予測値を算出できます。

「モデル選択」
グループ

木の数や変数の数などランダムフォレストのパラメータの選択を行うことができます

4. おわりに

機械学習は大量のデータから高い予測精度を持つ予測モデルを構築できるため、品質管理分野でも特にこれまでSQC（統計的品質管理）が扱えなかった領域において、機械学習によってより良い品質を確保できるようになることが期待されます^[5]。

一方、機械学習には次のような特徴があり、品質管理分野で機械学習を活用する際には注意が必要となります。

- ① 機械学習手法は予測を目的としている。一般的に機械学習は予測のプロセスがブラックボックス化される傾向にあり、多くの場合、機械学習は要因分析を目的とした分析では使用できない。
- ② 機械学習手法にはハイパーパラメータ（予測性能をチューニングするためのパラメータ）が存在し、ハイパーパラメータの設定によって得られる結果が大きく変わる。ハイパーパラメータは、扱っている問題に応じて適切に設定する必要があるが、決め方は難しい。

上記①に対しては、弊社は限定された状況の下では機械学習も要因分析で使用できると考えており、StatWorks/V5 機械学習編には状況によっては要因の分析でも使用可能な機械学習手法を搭載すると共に、各解析手法には要因分析で使用可能な出力をできるだけ用意しています。

また、上記②に対しては、StatWorks/V5 機械学習編では、できるだけユーザーがハイパーパラメータを適切に設定できるような機能構成や出力としています。

機械学習を品質管理で活用して成果を上げるためには多くの課題がありますが、StatWorks/V5 機械学習編がその課題の解決の一助となれば幸いです。

5. 参考文献

- [1]赤穂昭太郎(2008)：『カーネル多変量解析 非線形データ解析の新しい展開』，岩波書店
- [2]富岡亮太(2015)：『スパース性に基づく機械学習』，講談社
- [3]後藤正幸・小林学(2014)：『入門 パターン認識と機械学習』，コロナ社
- [4]井出剛・杉山将(2015)：『異常検知と変化検知』，講談社
- [5]小野田崇(2018)：“「製造業におけるビッグデータの解析あり方研究会」活動報告（第2報）
－SQCと機械学習－”，「品質」，48，[4]，12-15

本著作物は原著作者の許可を得て、株式会社日本科学技術研修所（以下弊社）が掲載しています。本著作物の著作権については、制作した原著作者に帰属します。

原著作者および弊社の許可なく営利・非営利・イントラネットを問わず、本著作物の複製・転用・販売等を禁止します。

所属および役職等は、公開当時のものです。

■公開資料ページ

弊社ウェブページで各種資料をご覧ください <http://www.i-juse.co.jp/statistics/jirei/>

■お問い合わせ先

(株)日科技研 数理事業部 パッケージサポート係 <http://www.i-juse.co.jp/statistics/support/contact.html>