

# 技術者のための StatWorks/V5 機械学習編の機能および有効な使い方

(株) 日本科学技術研修所  
数理事業部 犬伏 秀生

本発表では、弊社が 2019 年 4 月に発売を開始した機械学習ソフトウェア「JUSE-StatWorks/V5 機械学習編」(以下、機械学習編)の機能および有効な使い方をご紹介します。

※本内容は 2019 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. 機械学習編の概要・特徴

### (1) システムコンセプト

機械学習編は、主として製造業の各企業様が SQC (統計的品質管理) の中で機械学習を活用することを支援するための機械学習ソフトウェアです。このため、機械学習編は、専任のデータ解析担当者だけでなく、SQC に携わる多くの現場の技術者が手軽に機械学習を実行できるようになることを目指しています。

この目標を実現するため、機械学習編のシステムコンセプトは次のようになります。

《機械学習編のシステムコンセプト》

#### ① 計算処理も含め全て自社開発

機械学習編は、計算処理も含め、原則的に全て自社で開発しています。これにより、外部の制約を受けずにシステムを作成できています。また、各計算処理における定義式やアルゴリズムなども全て把握しているため、サポートの中でユーザー様にお伝えすることが可能です。

#### ② StatWorks/V5 シリーズのシステムコンセプトの引継ぎ

機械学習編は、StatWorks/V5 シリーズの既存製品 (総合編など) のシステムコンセプトを引き継ぎ、次の特徴を持つシステムとなります。

- ✓ 全てマウスで操作可能 (スクリプトの記述が不要)
- ✓ 分析結果をできるだけ視覚的に確認できるように豊富なグラフを用意
- ✓ 分析結果はグループ・タブに整理されて出力され、グループ・タブの並び順は解析ストーリーに則る

#### ③ 単体の PC 上で動作

機械学習編は、使っていただくためのシステム環境の敷居を下げるため、StatWorks/V5 シリーズの既存製品 (総合編など) と同様に単体の PC 上で動作するソフトウェアとしております。このため、機械学習編の動作にはサーバー、クラウドは不要です\*2。

※機械学習編ネットワーク版の場合はライセンス管理のためのサーバーが必要になります。

### (補足) 他の機械学習ツールとの比較

一般的に、市販の機械学習ツールは、専任のデータ解析担当者や情報システム担当者向けになっており、使用環境 (サーバー、クラウド上で稼働など) や操作性 (プログラミングが必要など) の面で、現場の技術者が手軽に使うには敷居が高い傾向にあります。

また、フリーソフトウェアである R、Python は、無料で使える、ライブラリが充実している、最新の分析手法を使用できる、Web 上や書籍で様々な情報を取得できる、などの優れた点があり、企業での利用も広がっています。しかし、R、Python を社内の機械学習の標準ツールとするには、次のよう

な問題があります。

《R・Python の問題点》

- ① スクリプトの記述（プログラミング）が必要となる。
- ② 動作上のトラブルが起きやすい。
- ③ 機械学習手法を学習するための教育で使用しづらい。
- ④ サポートの窓口が存在しない。

前述の通り、機械学習編は、これらの問題点をカバーしており、現場の技術者も含めた社内の幅広い人々が使う機械学習の標準ツールとしてご使用いただくのに適しています。

## （2）機能の特徴

機械学習編の機能上の特徴は次の通りです。

### ① StatWorks/V5 シリーズの既存製品（総合編 等）と一体で使用可能

スタンドアロン版、ネットワーク版共に、総合編などの StatWorks/V5 シリーズの既存製品と機械学習編を同じ PC にインストールすると、メニューが統合され、一体で使用することができます。

### ② ワークシートは StatWorks/V5 シリーズの既存製品と共通、機械学習手法の分析は専用のモジュール上で実行

データの inputs は、総合編などの StatWorks/V5 シリーズの既存製品と共通のワークシート上で行います。そして、機械学習手法のメニューを選択すると、専用のモジュールが起動します。機械学習の分析は、その専用モジュール上で行われます。

### ③ 解析可能なデータサイズは 1,000 変数(列) × 100,000 サンプル(行)以内

### ④ 稼働保証 OS は Windows10 (64bit)

## （3）有効な活用場面

機械学習編は、特に次の場面で有効にご活用いただけます。

### ① 中規模データの分析

機械学習編で分析可能なデータのサイズは 1,000 変数×100,000 サンプル以内ですが、現状では SQC で取り扱われるデータのサイズはこの範囲内であることが多いと認識しています。よって、機械学習編をご活用いただくことにより、SQC で取り扱われる多くのデータに対し、機械学習を用いた分析を手軽に行っていただけるようになります。

### ② 機械学習手法の教育

SQC の一環で機械学習を活用していくには、他の統計手法と同様に、社内外での教育を通じた人材育成が必要となります。特に社内で機械学習の教育を行う場合、前述のシステムコンセプトの通り機械学習編は、教育を実施する環境を構築し易い、演習を柔軟に実施できる等の特徴を持つため、機械学習編をご活用いただくと、教育をより効果的に行うことができます。

### ③ ビッグデータの分析モデルの検討

最終的にビッグデータを分析する場合でも、ビッグデータから抽出した中規模以下のデータを用いて、どの統計手法を使用すべきか、どの程度の予測精度になるか等の検討が必要になる場合があります。機械学習編は、一つのデータに対して様々な機械学習手法を手軽に実行でき、また、分析の設定を変更して再実行を行うことも簡単にできますので、このような検討に対しても機械学習編は有効です。

## 2. 機械学習編の搭載手法

### (1) 搭載手法一覧

機械学習編には、次の機械学習手法が搭載されています。

No	分類	解析手法
1	データクリーニング	データクリーニング
2	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



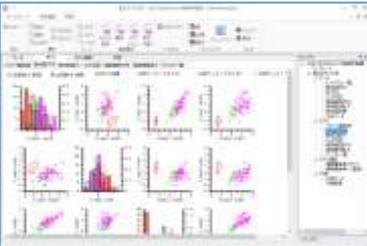
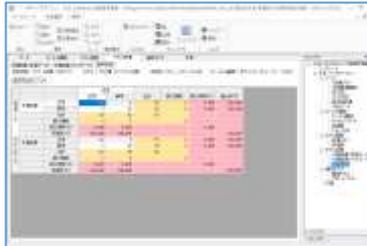
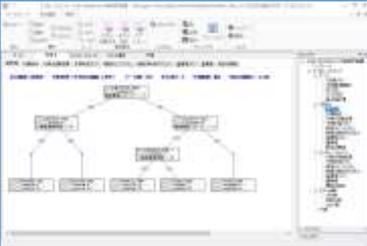
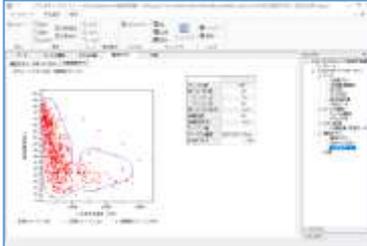
### (2) 多変量解析手法との対応

機械学習編に搭載した機械学習手法は、これまで SQC で使用されてきた多変量解析手法を拡張した統計手法と捉えることができます（機械学習手法と多変量解析手法との対応は渡邊[1]による）

No	機械学習手法(A)	対応する多変量解析手法(B)	B→A の主な拡張点	拡張のメリット
1	カーネル主成分分析	主成分分析	・ <u>カーネル法</u> の導入	複雑な構造のデータに対応
2	混合ガウス分布	階層的クラスター分析, 非階層的クラスター分析(k-means)	・ 正規分布の重ね合せを仮定	所属確率の取得
3	リッジ回帰, lasso 回帰, Elastic Net	重回帰分析・数量化 I 類	・ <u>正則化項 (罰則項)</u> の導入	多重共線性 (変数の数>サンプル数など) に対応
4	サポートベクターマシン (SVM)	判別分析・数量化 II 類	・ <u>カーネル法</u> の導入 ・ 分布の仮定の撤廃, 識別境界の決め方の変更 (マージン最大化) ・ <u>罰則項</u> の導入	汎化能力 (新たなデータに対する予測能力) の向上
5	ランダム・フォレスト	多段層別分析(AID), 多肢層別分析(CAID)	・ <u>バギング</u> の導入	汎化能力の向上
6	1クラス SVM	多変量管理図, MT 法	・ <u>カーネル法</u> の導入 ・ 分布の仮定の撤廃, 識別境界の決め方の変更	正規分布に従わないデータに対応
7	glasso	グラフィカルモデリング (GM)	・ <u>正則化項 (罰則項)</u> の導入	多重共線性に対応

### (3) 搭載手法の概要

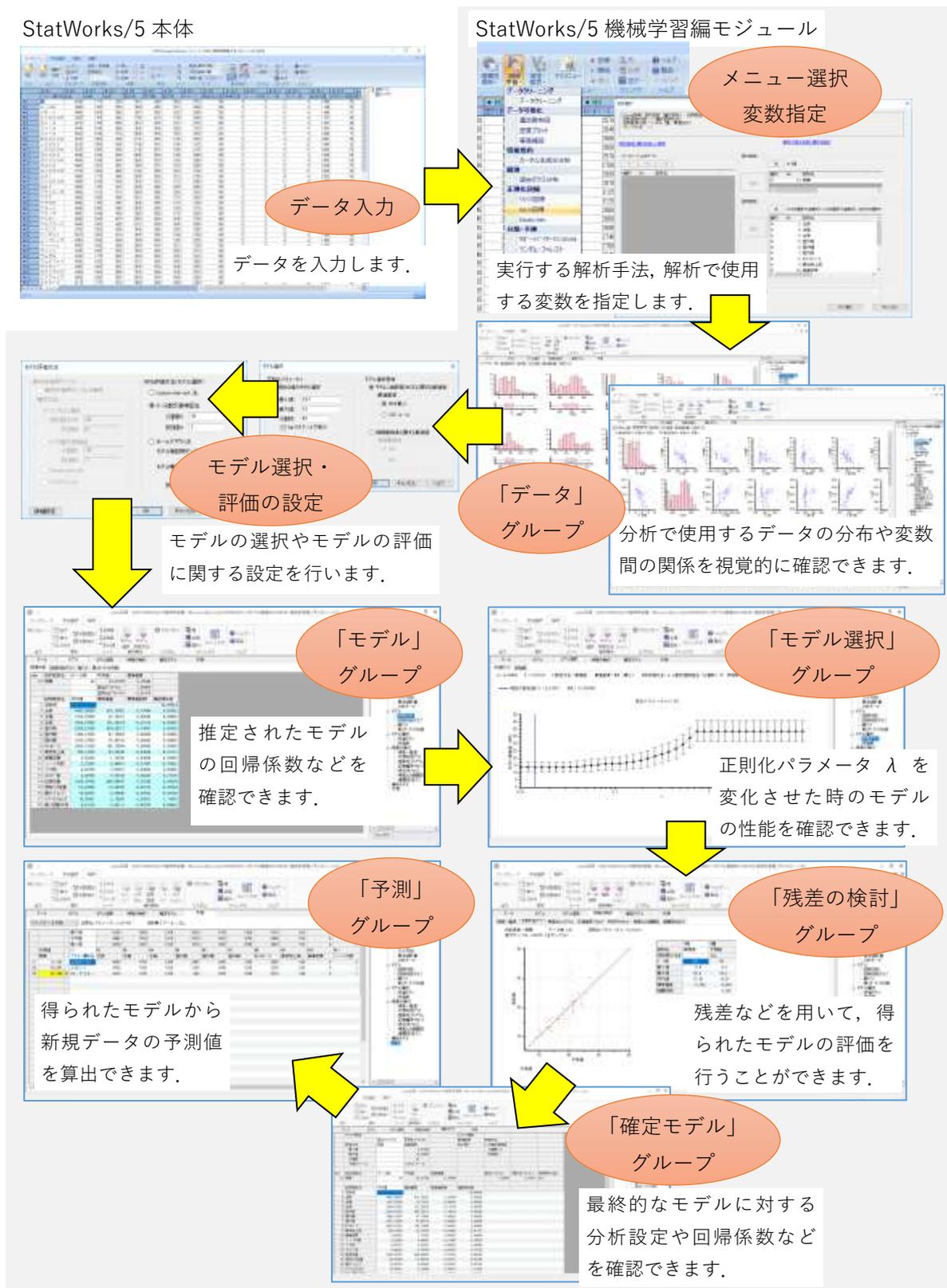
機械学習編に搭載されている機械学習手法の概要は次の通りです<sup>[2][3][4][5]</sup>。

<p>(3-1) 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</p> 	<p>(3-2) 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にデータの分布を視覚的に確認できます</p> 	<p>(3-3) カーネル主成分分析 データによっては, 主成分分析よりも低次元で良くサンプルを分類できる場合があります。</p> 
<p>(3-4) 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し, それに基づいてサンプルの層別を行うことができます</p> 	<p>(3-5) 正則化回帰 (リッジ回帰・lasso 回帰・Elastic Net) 説明変数の数がサンプル数よりも多い状況でも回帰式を当てはめて精度良い予測を行えます。</p> 	<p>(3-6) サポートベクターマシン(SVM) カーネル法やマージン最大化による分類境界線の決定等により目的変数(質的)をより良い精度で予測することができます。</p> 
<p>(3-7) 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。</p> 	<p>(3-8) 1クラス SVM 正常データのみから外れ値を検出するための識別境界を構築することができます。</p> 	<p>(3-9) 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。</p> 

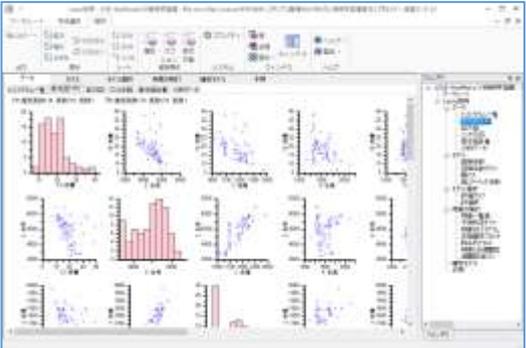
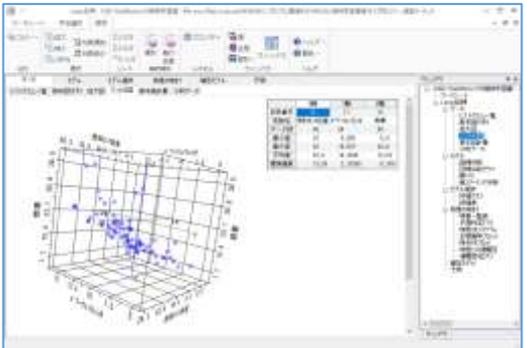
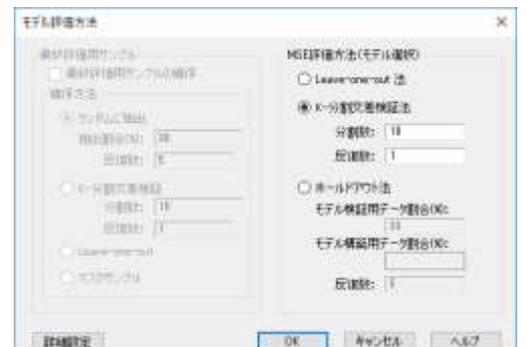
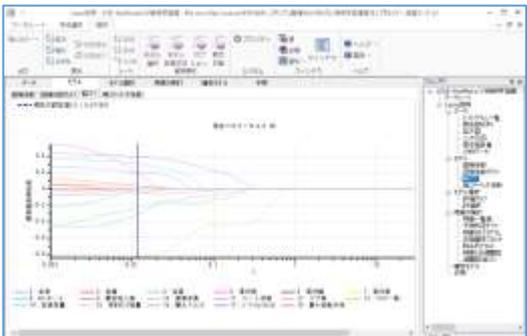
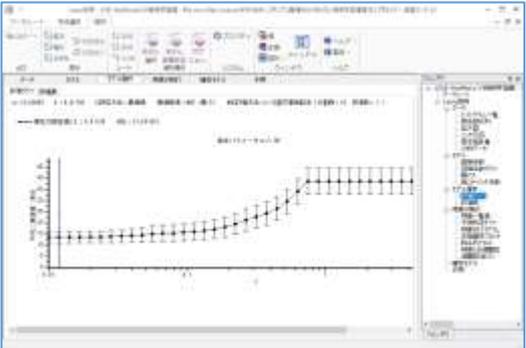
### 3. 機械学習編の分析手順・出力画面

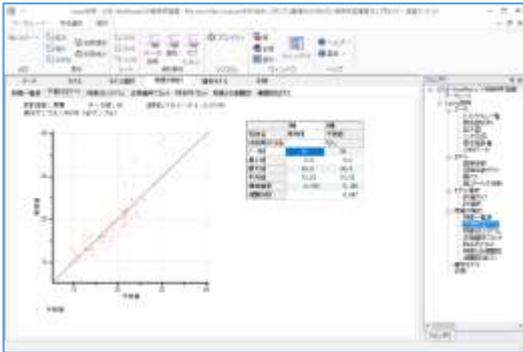
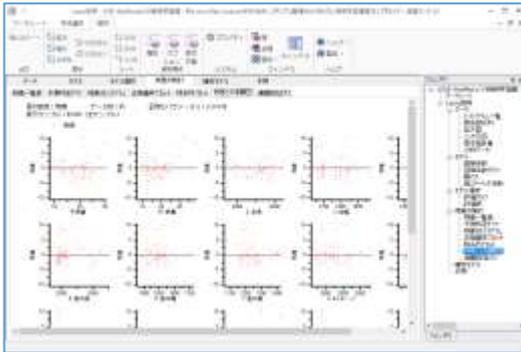
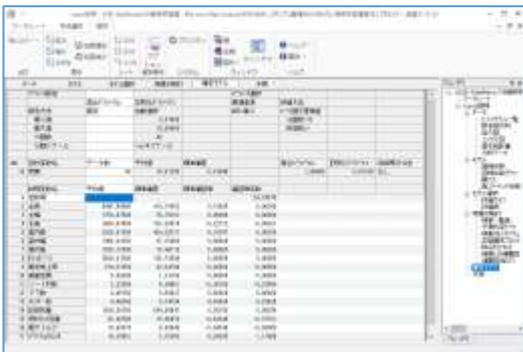
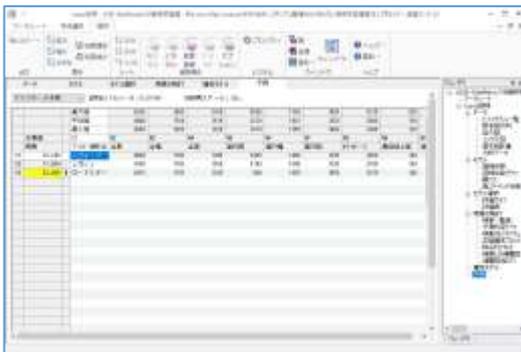
ここでは、解析手法「lasso 回帰」を例に、機械学習編の分析手順、出力画面を紹介します。

#### (1) 解析手法「lasso 回帰」の分析手順



(2) 解析手法「lasso 回帰」の主な出力画面

データ	<p>(a) 散布図行列</p> 	<p>(b) 三次元図</p> 
分析設定	<p>(a) 「モデル選択」ダイアログ</p> 	<p>(b) 「モデル評価方法」ダイアログ</p> 
モデル	<p>(a) 回帰係数</p> 	<p>(b) 解パス</p> 
モデル選択	<p>(a) 評価グラフ</p> 	

残差の検討	(a) 予測判定グラフ 	(b) 残差との連関図 
確定モデル ・予測	(a) 確定モデル 	(b) 予測 

#### 4. おわりに

本発表では、弊社が今年発売を開始した機械学習ソフトウェア「JUSE-StatWorks/V5 機械学習編」の機能や有効な使い方のご紹介を行いました。

機械学習編が、皆様が機械学習を活用し、成果を上げる一助となれば幸いです。

#### 5. 参考文献

- [1]渡邊克彦, 品質・技術力向上に繋げる SQC と機械学習のよりよい使い方について,  
日本科学技術研修所, 第 29 回 JUSE パッケージ活用事例シンポジウム要旨集(2019)
- [2]赤穂昭太郎, “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, 2008
- [3]富岡亮太, “スパース性に基づく機械学習”, 講談社, 2015
- [4]後藤正幸・小林学, “入門 パターン認識と機械学習”, コロナ社, 2014
- [5]井出剛・杉山将, “異常検知と変化検知”, 講談社, 2015

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>