SAS Viya Data Mining and Machine Learningプレビュー

宮崎　洋
（SAS Institute Japan株式会社）

SAS Viya Data Mining and Machine Learning preview

Hiroshi Miyazaki
SAS Institute Japan

要旨 :

SASの新しいアーキテクチャであるSAS Viya製品の一つであるSAS Viya Data Mining and Machine Learning(略称 VDMML)の機能をリリース前のプレビュー版としてSASプログラミングを通して紹介する。

キーワード : SAS Viya,SAS Studio, Data Mining and Machine Learning

アジェンダ

1. SAS Viya Data Mining and Machine Learning の概要
2. システムの概要
3. SAS Viya Data Mining and Machine Learningの実装方法

1.SAS Viya Data Mining and Machine Learning の概要

SAS Viya Data Mining and Machine Learning（略称：VDMML ）は、データ探索、ビジュアライゼーション（視覚化）、さらには統計、データマイニング、機械学習の最新手法の全てを、SAS StudioのWebインタフェース上で実行可能であり、インメモリ処理の環境化において、高速な処理が可能です。

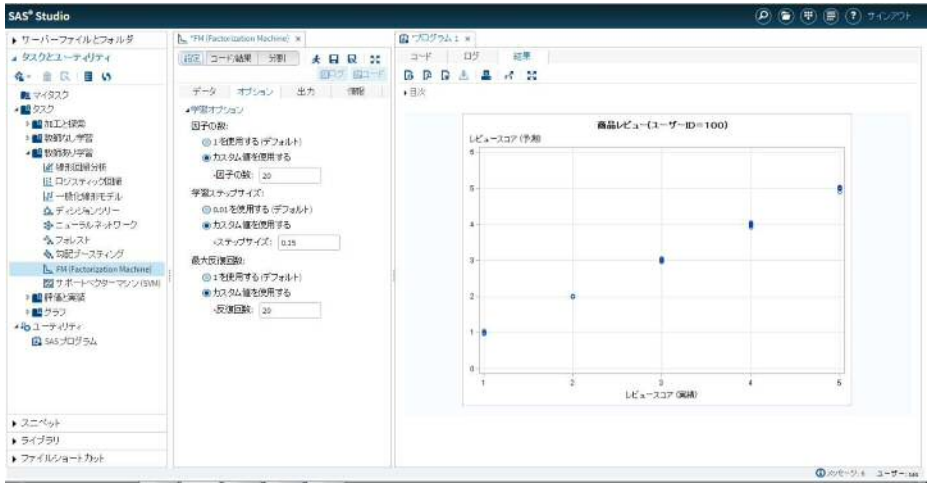
主な特徴

・Webベース、対話操作型の柔軟な分散インメモリ・プログラミング環境を提供

・サーバーサイドで集中管理が可能

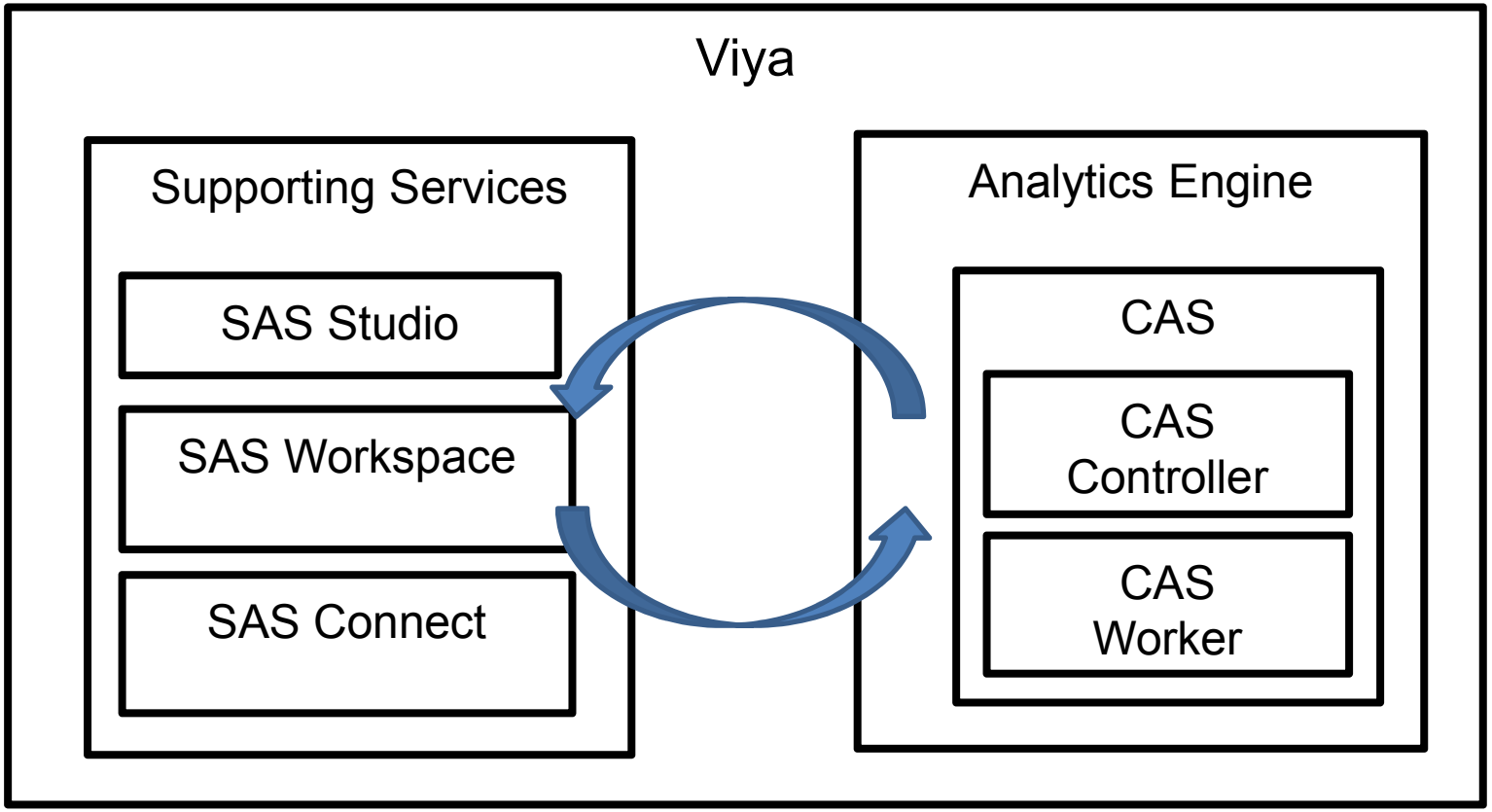
・機械学習タスクを迅速に開始、自動化するためのSASコード自動生成

・計算処理に最適化された、次世代のSAS In-Memory Analytics の利点の活用　等



2.システムの概要(SAS Viyaのアーキテクチャー)

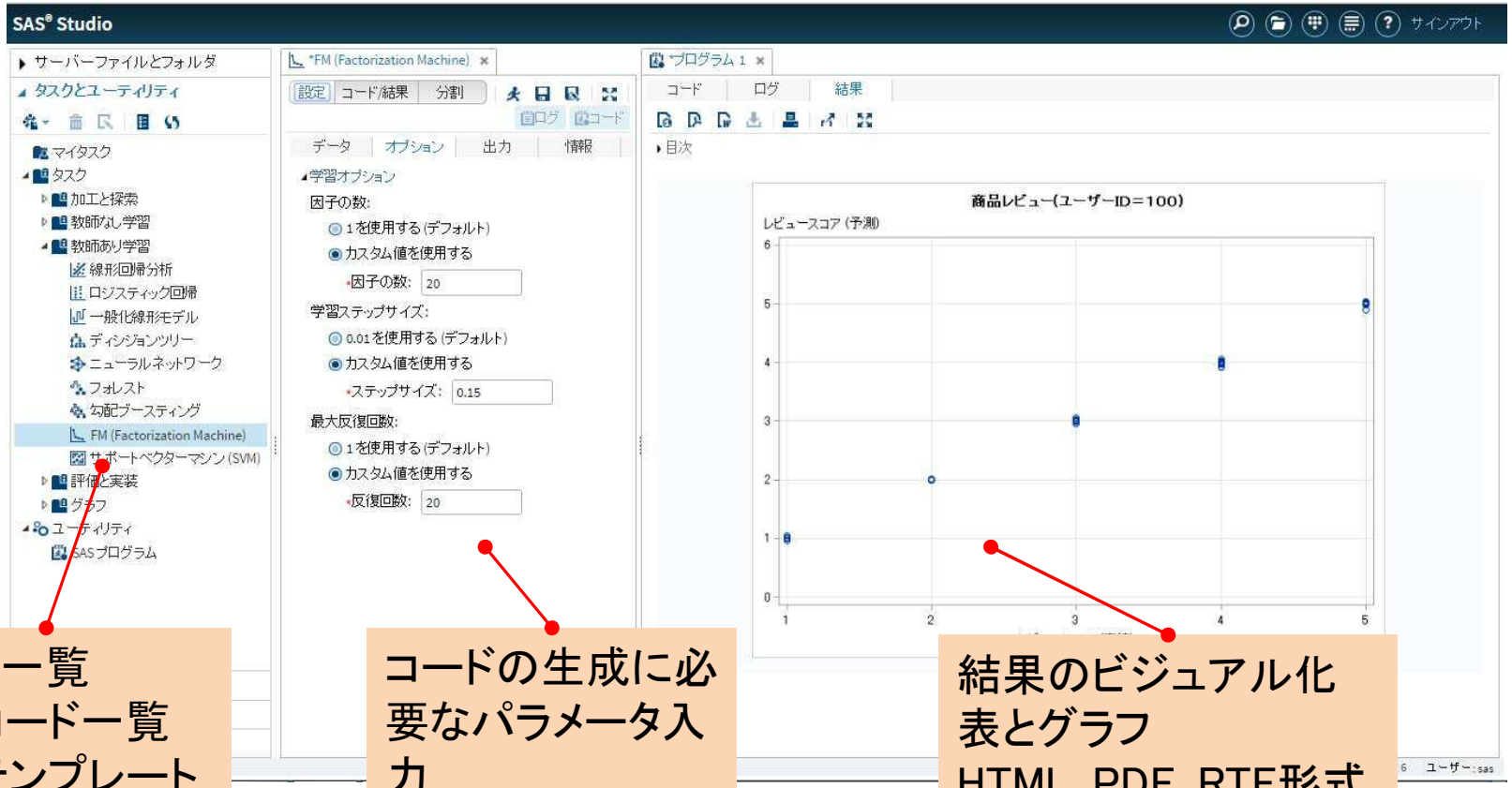
SAS Viyaは、SAS言語を実行する環境である、SAS Studioを利用します。開発などに必要なデータは、ローカルやCAS上へデータをアップロードして、探索や分析を行います。



SMP版での実行アーキテクチャ

2.システムの概要（SAS Studioの概要）

Web ベースのプログラミング環境を提供し、一般的な機械学習のタスクを選定することが可能です。また、パラメータ入力により、バッチ実行や自動化で利用できるSASコードが自動生成されます。



2.システムの概要(VDMMLの主な機能)

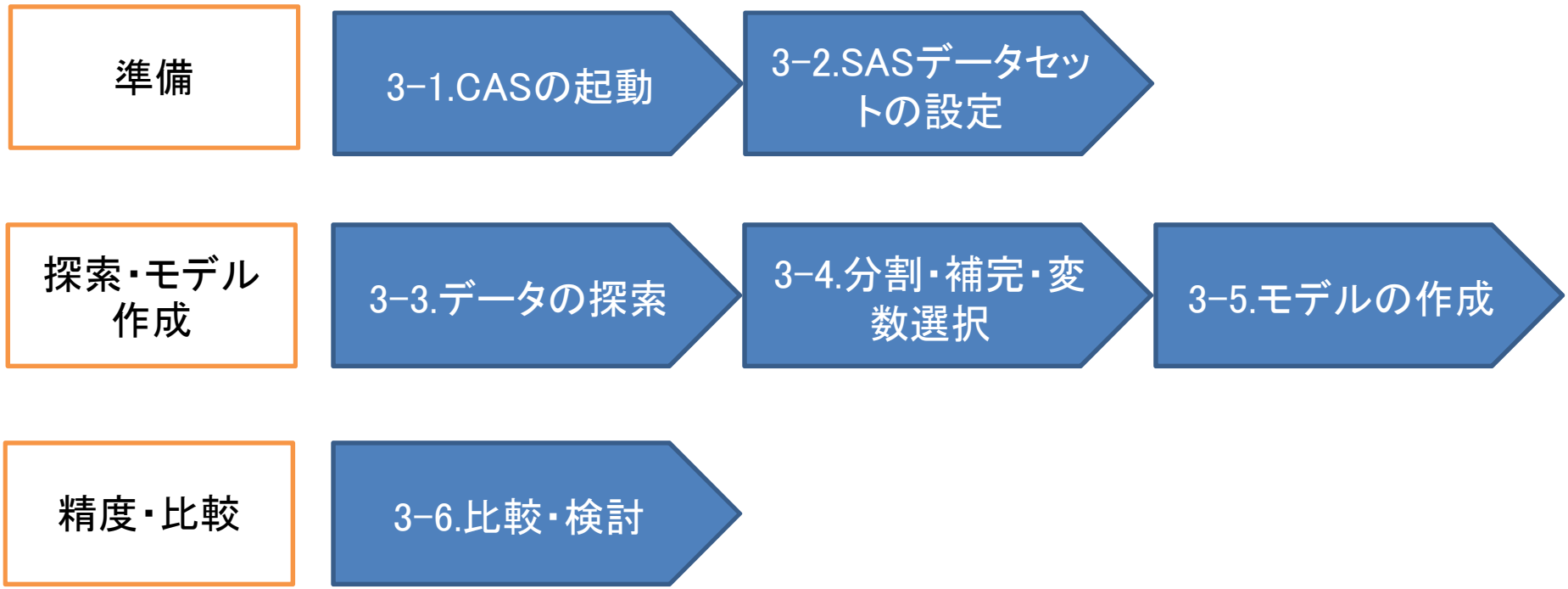
VDMMLは、以下の主な機能が利用可能となります。

SAS Visual Data Mining and Machine Learning	
FACTMAC	ファクタライゼーションマシーン
FOREST	ランダムフォレスト
GRADBOOST	勾配ブースティング
NNET	ニューラルネットワーク
SVMACHINE	サポートベクターマシン
TEXTMINE	テキストマイニング
TMSCORE	ドキュメントスコアリング

Common Analytical Procedures	
ASSESS	アセスメント
BINNING	ビン化
CARDINALITY	要約
PARTITION	サンプリングと分割
VARIMPUTE	補完
VARREDUCE	変数選択

3.SAS Viya Data Mining and Machine Learningのプログラミング実装方法

SAS Viya Data Mining and Machine Learningの実装は、SAS Studioをインターフェースとして、プログラミングを行います。今回は、以下の流れに従い実装をします。



3-1.CASの起動

最初のステップは、CASのセッションを作成し、CASセッションに接続するために使用できるCASエンジンを設定します。

```
cas {casセッション名}

libname {casセッション名} sasioca sessref={casセッション名};

caslib _all_ list sessref={casセッション名};

 run;
```

・CASのセッションを作成し、「cas」ステートメントで任意のセッション名を指定します。
・libname ステートメントを使用してCASセッションをライブラリとして登録します。
・ caslib ステートメントを実行し、セッションの情報を出力します。

3-2.SASデータセットの設定

CASセッションのアクセス後、事前に配置している、SASデータセットをロードします。

```
data {CAS上のデータセット}
set  {ファイルシステム上のデータセット}
run;
```

・libname ステートメントで登録した、CASライブラリに対してデータをロードします。
・事前にローカルのライブラリには、SASデータセットがあり、対象のデータセットをアップします。

補足:

CASテーブルの持ち方

・グローバルor セッション単位
・ 一時的　or 永続的
・メモリ or ディスク
・重複 (Table A)　or 分散(tableB)

3-3.データの探索

CAS上のテーブルを使用し、分析に関するデータを調べます。PROC CARDINALITYを実行すると、変数定義、基本的な統計情報、および欠損値などの統計情報を出力します。

```
proc cardinality data={CAS上のデータセット}
outcard={CAS上のデータセット};
・・・・
```

PROC CADINALITY の実行によって、以下のテーブルが作成されます。

・カーディナリティのデータテーブル
・詳細データテーブル

3-4.分割・補完・変数選択

分割・補完・変数選択を実施し、次のステップのモデルを比較する為のデータ加工を行います。

	プロシージャー	主な機能
サンプリング・データ分割	proc partition data={CAS上のデータセット} samppct= {サンプリング}; By {変数選択};	・入力データの単純ランダムサンプリング、層別ランダムサンプリング、オーバーサンプリングを実行
欠損補完	proc varimpute data={CAS上のデータセット} Input {変数名} /ctech {置き換える値};	・欠損値を平均値、中央値、乱数等で置き換える等
変数選択	proc varreduce data={CAS上のデータセット} technique= {役割} class={ターゲット変数} ;	・ターゲット変数,説明変数のセットを特定することで、教師ありまたは教師なしの変数選択などを実行

3-5.モデルの作成

3つの予測モデルの機能を利用し、最適なモデルを検討します。

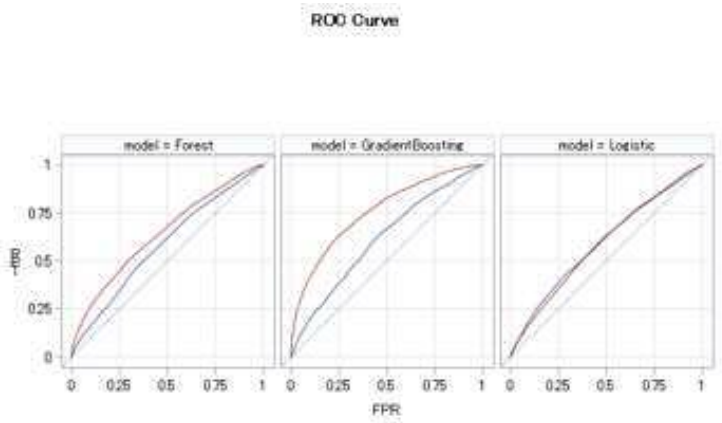
	プロシージャー	主な機能
ロジスティック回帰	proc logselect data={CAS上のデータセット}; class {クラスに設定する変数}.;	・ロジスティック回帰タスクは、二項応答モデルのロジスティック回帰と自動モデル選択を使用して予測モデルを当てはめます。
ランダムフォレスト	proc forest data={CAS上のデータセット} ntrees={反復回数} intervalbins={ビン間隔} minleafsize={葉のオブザベーション数};	・ランダムフォレスト法を使用して、ディシジョンツリーの集合を作成する ・間隔尺度または名義尺度のターゲット変数の予測モデルを作成します。
勾配ブースティング	proc gradboost data={CAS上のデータセット} ntrees={反復回数} intervalbins={ビン間隔} maxdepth=5;	・勾配ブースティング法を使用して、ディシジョンツリーの集合を作成する間隔尺度または名義尺度のターゲット変数の予測モデルを作成します

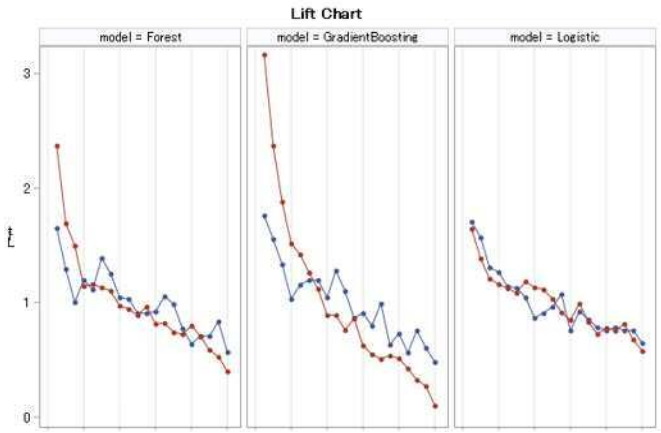
3-6.比較・検討

PROC ASSESSを使いスコアリングされたデータテーブルに基づきモデルの評価を行います。また、SGPLOTによりROCチャート等でグラフィカルに把握することが可能です。

```
proc assess data={スコアリングされたデータセット};
  input {予測ターゲット};
  target  {ターゲット} ;
・・・・・
```

教師付き学習モデルの評価の実施
- リフトチャート
- ROCチャート





参考資料

• SAS Studio Tutorials
http://support.sas.com/training/tutorial/studio/

• SAS Studio 3.4 User's Guide
http://support.sas.com/documentation/cdl_alternate/ja/webeditorug/68254/PDF/default/webeditorug.pdf

・SAS Japan Blog (すべてのSASユーザーのためのSAS® Studio)
http://www.sascom.jp/blog/2015/08/12/sas-studio/

・SAS® Viya™の全体紹介
http://www.sas.com/ja_jp/software/viya.html

• SAS® Viya™の概要資料
http://www.sas.com/content/dam/SAS/ja_jp/doc/factsheet/so-sas-viya-108233-jp.pdf

• SAS® Viya™ Data Mining and Machine Learningの全体紹介
http://www.sas.com/ja_jp/software/analytics/data-mining-machine-learning.html

• SAS® Viya™ Data Mining and Machine Learningの製品資料
http://www.sas.com/content/dam/SAS/ja_jp/doc/factsheet/fs-sas-viya-data-mining-machine-learning-108275-1606.pdf

・SAS® Viya™ Data Mining and Machine Learning trial
http://www.sas.com/en_us/trials/try-data-mining-machine-learning/ea-form.html

ご清聴ありがとうございます