

SAS Technical News

Winter 2001

*For Higher
Customer Satisfaction,
We Bridge
the SAS System
Between
Customer's World.*

CONTENTS

- 1 特集 Enterprise Miner ソフトウェア バージョン4.0のご紹介
~バージョン 2.0xからの拡張点を中心に~
- 7 Q&A
- 12 SASトレーニングのお知らせ
- 12 新刊マニュアルのご紹介

特集

Enterprise Miner ソフトウェアバージョン4.0の ご紹介

~バージョン 2.0xからの拡張点を中心に~

はじめに

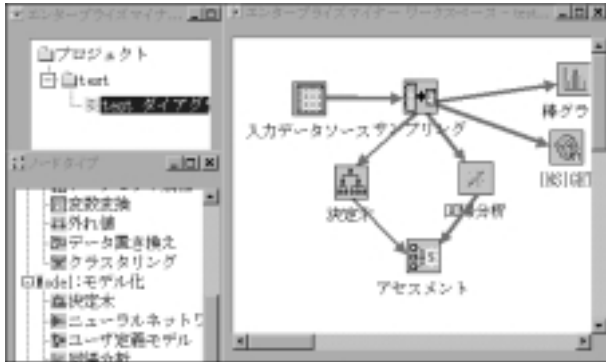
この特集では、はじめにデータマイニングについての簡単な説明を行い、次にEnterprise Miner ソフトウェア バージョン4.0 (以下EM4) において新規に追加された機能、及びEnterprise Miner ソフトウェア バージョン2.0x(以下EM2)からの主な拡張点をご紹介します。

1. データマイニングとは? ~EM2における機能~
2. インターフェイスの変更
3. ターゲットプロファイルについて
4. 新規に追加されたノードの紹介
5. 既存のノードについて
6. その他

1. データマイニングとは? ~EM2における機能~

データマイニングとは、大規模なデータからそれまでは見つけだすことのできなかつた有益な情報を導き出すための一連のプロセスであると言えます。具体的には、分析データを準備し、データの大まかな分布を調べて、変数変換、クラスタリングなどの作業を施してから、回帰分析やニューラルネットワークの手法を適用し、実行結果の有用性と信頼性を評価するという流れになります。Enterprise Minerソフトウェアでは、この流れを、「サンプリング」・「探索」・「加工」・「モデル化」・「評価」の5つの単位に分類しています。EM2では、「ニューラルネットワーク」、「決定木」、「回帰分析」などのデータマイニングで使用される分析手法が用意されています。また、「ノード」と呼ばれるアイコンをワークスペースにドラッグ&ドロップしてそれらをつなげるだけで、プログラミングなしで分析できるといったユーザーフレンドリーなGUIを実現しています。このように、ユーザーは容易に分析手法をカスタマイズすることができます。

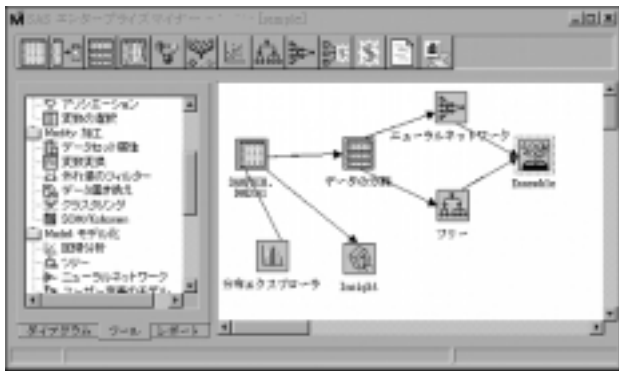
EM2のウィンドウ例)



[プロジェクトを管理するプロジェクトウィンドウ(左上)、ノードを表示するノードウィンドウ(左下)ワークスペースウィンドウ(右)]

2. インターフェイスの変更

EM2では、プロジェクト、ノードタイプ、ワークスペースの各ウィンドウが別個に存在していましたが、EM4のインターフェイスでは、全てが一つのウィンドウに表示されるようになりました。これにより、プロジェクト管理が容易になりました。



ウィンドウの左側はプロジェクトナビゲータと呼ばれる領域で、プロジェクトを管理する「ダイアグラムタブ」、各ノードを表示する「ツールタブ」、及びレポートノードによって出力されるレポートを管理する「レポートタブ」があり、それらを切り替えて表示することができます。

また、頻繁に使われるノードは、上部にツールバーとして並べられており、そこからノードをワークスペースに直接ドラッグ&ドロップすることも可能です。

3. ターゲットプロファイルについて

EM4日本語版で追加された機能の一つに、ターゲットプロファイルがあります。EM4では、ある意思決定を行った時に生じるコストや利益情報をターゲットプロファイルと呼んでいます。これにより、ユーザーは費用を最小化もしくは利益を最大化するような意思決定を選択することができ、コスト意識をもったデータマイニングが可能となります。

具体例を考えてみましょう。顧客がある商品を購入した場合YES、しなかった場合NOが変数PURCHASEの値であるとして。また、顧客にカタログを送っていた場合に10,000円分の商品を買うものと仮定し、さらに、カタログを送るためのコストが500円であるとします。ある顧客が商品を購入する確率を0.30、購入しない確率を0.70と仮定すると、次のような利益の予測が成り立ちます。

	カタログを送った	カタログを送らなかった
YES	$10000 \times 0.30 = 3000$	$10000 \times 0.00 = 0$
NO	$0 \times 0.70 = 0$	$0 \times 1.00 = 0$
合計	3000	0
コスト	-500	0
利益	2500	0

この顧客にカタログを送付した場合、2,500円の利益が期待されます。

別の顧客について考えてみましょう。商品を購入する確率を0.02、購入しない確率を0.98と仮定すると、次のような予測が成立します。

	カタログを送った	カタログを送らなかった
YES	$10000 \times 0.02 = 200$	$10000 \times 0.00 = 0$
NO	$0 \times 0.98 = 0$	$0 \times 1.00 = 0$
合計	200	0
コスト	-500	0
利益	-300	0

この場合には、カタログを送ったときの期待値は300円の赤字です。このような設定をターゲットプロファイルで定めて分析することにより、いかなる販売戦略を選択するべきかを調べることができます。ターゲットプロファイルは、以下の各ノード上で定義、編集、または表示することができます。

- ・ 入力データソース
- ・ データセット属性
- ・ 回帰分析
- ・ ツリー
- ・ ニューラルネットワーク
- ・ アンサンブル
- ・ ユーザー定義のモデル

また、データセットのターゲットプロファイルを作成して、プロジェクトレベルでターゲットプロファイル全てに対するデフォルト設定を指定することができます。

4. 新規に追加されたノードの紹介

EM4では、新たに次の4つのノードが用意されました。それらを順にご紹介いたします。また、この章と次の第5章で記述するノード名の横の括弧内には、冒頭の「1.データマイニングとは？」でご説明したノードの役割です。

- ・ レポーターノード
- ・ SOM/Kohonenノード
- ・ アンサンブルノード
- ・ マルチプロットノード

4.1 レポーターノード(評価)

レポーターノードを接続することにより、それまでのマイニングのプロセスと結果の要約を、HTML形式で出力できるようになりました。

レポートの出力例)

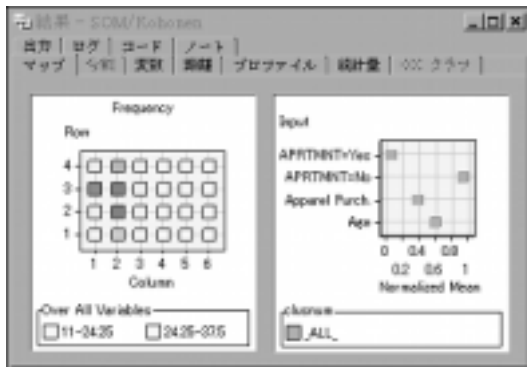


レポートの中でリンクされている箇所をクリックすると、それらに関するより詳細な情報が表示されます。

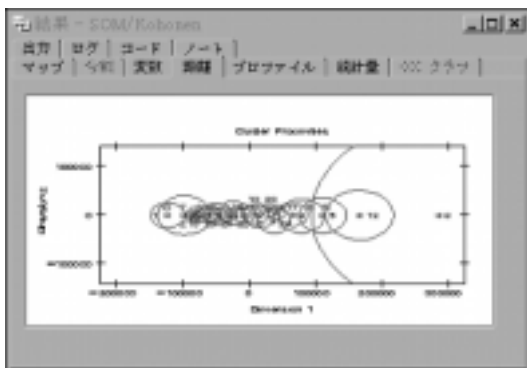
4.2 SOM/Kohonen ノード (加工)

SOM/Kohonen ノードは、近年盛んに研究されている「自己組織化マップ(Self Organizing Maps)」や、「ベクトル量子化 (Vector Quantization)」と呼ばれる手法を用いて、オブザベーションをクラスター化、またはセグメント化します。SOMでは、高次元のデータ空間を、人間が視認できる2次元で再現させることが可能です。オプションの指定により、分析の際に用いる設定値、例えばマップのサイズ、近傍のサイズやクラスターの数などを柔軟に変更することができます。また、解析結果は、様々なグラフの形で出力されます。

自己組織化マップの実行情例)



[4x6の格子を用いた自己組織化マップの表示(左)入力変数の平均のグラフ表示(右)]



[各クラスターのサイズ、およびクラスター間の距離、関係の表示]

自己組織化マップやベクトル量子化など、手法の詳細については、(下記)をはじめとする各種専門書籍で紹介されていますのでそちらをご参照ください。

自己組織化マップ

T. コホネン著 シュプリンガー・フェアラーク東京

(T.Kohonen: Self-Organizing Maps Springer 1995の日本語訳)

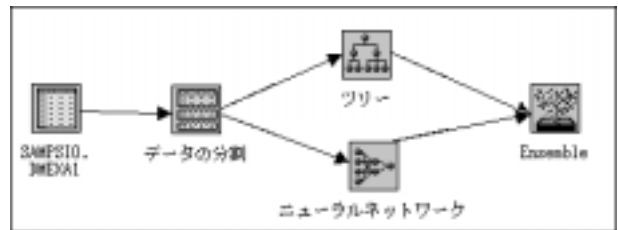
4.3 アンサンブルノード (モデル化)

アンサンブルノードは、複数のモデルから得られた予測値の「平均」を求めめるためのノードです。その後、この新しいモデルを使用して、新しいデータをスコアすることができます。通常は、以下のような形態で用います。

・ Combined model (結合モデル)

複数のモデルから予測された値を平均化することにより作成されます。結合モデルは、ニューラルネットワークノード・ツリーノードなどによる予測値の平均を新しい予測値とします。場合によっては、単一モデルだけを用いるよりも予測の精度が向上します。

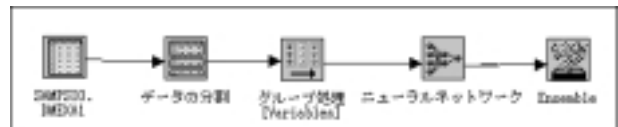
[ダイアグラム]の例)



・ Stratified model (層別モデル)

学習データを複数のグループに分け、それぞれに対してモデルを生成し、予測値を求めます。層別モデルは、同一データに複数の母集団が存在する場合に役立ちます。

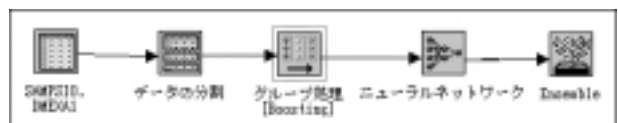
例)



・ Bagging/Boosting models

baggingモデルおよびboostingモデルは、学習用データから重複を許したサンプリングを行い、元データの「コピー」を幾つか作成します。そして、そのコピーデータごとにモデルを当てはめるといふ処理を行います。その後、各予測値の平均を新しい予測値とします。baggingやboostingについては、後段の「グループ処理ノード」の項もご参照ください。

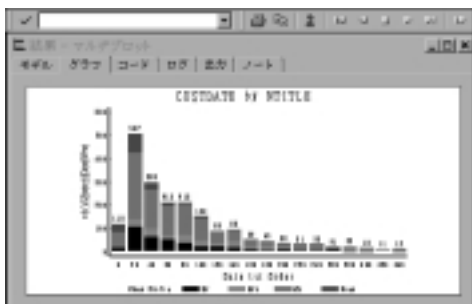
例)



4.4 マルチプロットノード (探索)

マルチプロットノードは、分布のグラフを自動的に作成して表示します。また各変数とターゲット変数との関係をグラフにしたり、散布図や回帰式の算出結果も表示できます。

マルチプロットノードによる出力例)



[上部ツールアイコンの一番右のボタンを押すと、各グラフがスライド表示されます。]

5. 既存のノードについて

EM4では、既存のノードに対しても多岐に渡る拡張が行われました。ここでは、EM4における主な拡張点をご紹介します。

5.1 グループ処理ノード (ユーティリティ)

グループ処理ノードを使うと、1つの処理フロー内で複数のターゲット変数を分析できます。また、グループの各水準ごとの処理も行えます。グループ処理のモードが以下の6つの中から選べるようになりました。

- ・ Variables
EM2のグループ処理ノードと同じ処理を行います。この「Variables」がグループ処理ノードにおけるデフォルトのモードです。
- ・ Cross-Validation[Variables] (交差検証法)
指定したグループについて、それらを含まないグループを分析の対象とします。

例えば、グループ変数GENDERが2水準M, Fを持っているものとし、またグループ変数REGIONが4水準N, S, E, Wを持っていると仮定します。この場合、「Variables」をモードとして選ぶと、グループ処理ノードでは次のように8回のループが行われます。

Loop	GENDER	REGION
1	M	N
2	F	N
3	M	S
⋮	⋮	⋮
8	F	W

一方、「Cross-Validation(Variables)」を選択すると次の8回のループが行われます。

Loop	
1	(GENDER=M REGION=N) ではないグループ
2	(GENDER=F REGION=N) ではないグループ
3	(GENDER=M REGION=S) ではないグループ
⋮	⋮
8	(GENDER=F REGION=W) ではないグループ

- ・ Unweighted resampling for bagging
重複を許したりサンプリングを行います。これは、単純無作為抽出によってデータの「コピー」をいくつも作成したい場合に用います。アンサンブルノードと組み合わせて、baggingを行うことができます。baggingを行うと、ツリー分析などの予測精度の向上が期待されます。
- ・ Weight resampling for boosting
この手法も、重複を許したりサンプリングを行います。ただし、抽出の際に、誤分類されたレコードを抽出する確率が高くなります。
- ・ Index
全く同じデータを後続の処理に渡します。自分自身でカスタマイズしたりサンプリング方法を設定したいときに用います。
- ・ Noloop
ループ処理を行いません。通常、ループ処理には計算時間がかかりますので、テスト的に処理を行うときに用いてください。

5.2 変数変換ノード (加工)

EM4では、4つの変数変換の手法が追加されました。

- ・ Optimal binning for relationship to target transformation
ツリー分析によって、間隔変数の値を幾つかのグループに分割します(ターゲット変数が2値の場合に限る)。
- ・ Maximize normality (正規性の最大化)
- ・ Maximize correlation with target(ターゲットとの相関の最大化)
- ・ Equalize spread with target levels(ターゲット水準との分布の均等化)
上記3つの変数変換の手法は、まず元となる変数Xに対して以下のつの変数変換を行い、次にそれぞれの基準に一番適した変換を選択します。

```
X (変換なし) log X, X0.25, X0.5, X2, X4
```

Maximize normality
6つの変数変換の中で、変換後の分布の歪度が最も0に近いものを選び、それを結果として返します。

Maximize correlation with target
ターゲット変数との相関係数の絶対値が一番大きいものを選択します。なお、ターゲット変数が間隔変数でない場合には、この手法を用いることができません。

Equalize spread with target levels
ターゲット変数の各水準において、間隔変数の分散を安定化させる方法です。

5.3 データ置き換えノード (加工)

欠損値や外れ値などを、ユーザー指定の方法で置き換えるノードです。間隔変数、分類変数両方に対する新たな置き換え手法として、以下の各手法が追加されました。

- ・ Distribution based(分布ベース)
データから求められたパーセント点を、一様乱数に従って割り当てます。
- ・ Tree imputation(ツリー分析による置き換え)
ツリー分析によって、欠損値を埋めます。他の入力変数によってツリー分析を行い、その予測値で欠損値を置き換えます。
- ・ Tree imputation with surrogates (代理ルールを伴うツリー分析による置き換え)
この手法も、ツリー分析によって欠損値を埋めます。ただし、ツリー分析の際に代理ルール(surrogate rule)を用います。

- ・ Default constant (デフォルト定数)
与えた定数で置き換えを行います。

また、間隔変数のみに適用できる手法として、次の4つも利用できるようになりました。

- ・ Mid-minimum spacing (最小区間の中心)
データの中で一番密度が高い区間を探して、その区間の「中心」の値を欠損値へ代入します。まず、データがN%残るような、また残ったデータの範囲が最小になるような区間を求めます。具体的には、その区間の上限と下限を加えて2で割った値を、欠損値へ代入します。また、Nはユーザーによって指定された値です。

- ・ Tukey's biweight (Tukeyの双加重)
- ・ Hubers (Huberの関数)
- ・ Andrew's Wave (Andrewの波形)

欠損値をこれらの推定値で埋めます。これらは、位置パラメータに対するM推定量です。平均などの推定量は、外れ値の影響を強く受けます。M推定量は平均などに比べ、より外れ値に対して抵抗性があります。

5.4 クラスタリングノード (加工)

クラスター分析を実行し、データをいくつかのセグメントに分けるノードです。Cubic clustering criterionに基づいて、クラスター数を決定することができるようになりました。

5.5 回帰分析ノード (モデル化)

回帰分析を行うノードです。EM4では、変数間の散布図も出力できるようになりました。これにより、残差の分布などをグラフィカルに表示することが可能です。

5.6 ツリーノード (モデル化)

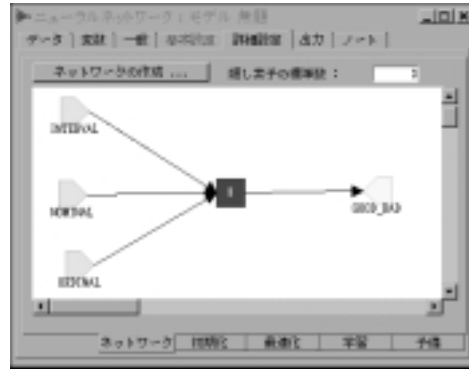
ツリー分析を行うノードです。IF-THENルールに基づいて、データを分割、分類します。EM2の「決定木ノード」から名称が変更され、また次のような拡張が行われました。

- ・ ターゲット変数として、順序変数も利用できるようになりました。
- ・ 剪定 (pruning) を行う際の基準が追加されました。上位10、25、50%の平均利益が最大になるように、作成された木を剪定することができるようになりました。
- ・ メニューから「ファイル」>「ルールの保存」で、外部テキストファイルにルールを保存できるようになりました。

5.7 ニューラルネットワークノード (モデル化)

ニューラルネットワークは、人間の脳の神経回路網に似たメカニズムを作りだそうとする研究者たちによって開発されたものです。近年、こうした手法に統計学と数値解析の方法が取れ入れられるようになりました。ニューラルネットワークは、データ内の複雑な非線形関係を検出するのに役立ちます。EM4では、ネットワークの構成を「基本設定」、「詳細設定」のいずれかから行うことができるようになりました。デフォルトで設定されている「基本設定」では、あらかじめいくつかのテンプレートが用意されており、より簡単にネットワークの構成を行うことができます。一方、「一般」タブの「ネットワークダイアグラムから指定する」を選択すると「詳細設定」タブが有効になり、こちらではネットワークフローをユーザーが自由に設定することができます。

ネットワークフローのカスタマイズウィンドウ例)



5.8 SASコードノード (ユーティリティ)

Enterprise Minerソフトウェア上でSASプログラムを用いることができます。このノードには、データセットと変数を動的に参照するためのマクロ機能が用意されています。EM4では、「マクロ」タブにおいて、マクロ変数を階層的に表示、管理できるようになりました。また、扱うことのできるマクロ変数が大幅に増えました。

SASコードノードを開いたときに表示されるウィンドウ例)



5.9 サンプリングノード (サンプリング)

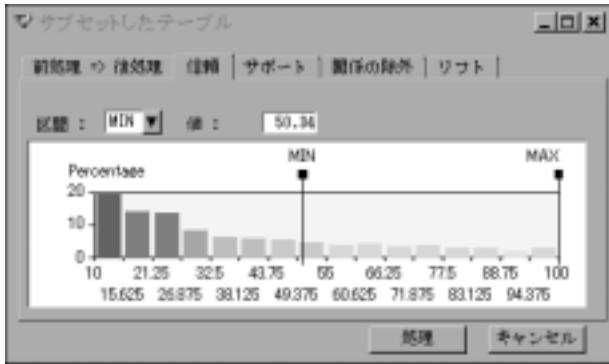
サンプリングノードは、膨大なデータからサンプリングデータを抽出する役割を果たします。EM4では、ユーザー指定の層別サンプリングがサポートされました。

5.10 分布エクスプローラノード (探索)

分布エクスプローラノードを用いると、2次元や3次元のグラフを描くことができます。EM2では棒グラフノードと呼ばれていましたが、EM4でこの名前に変更されました。また、グラフの設定方法がより柔軟になりました。

5.11 アソシエーションノード (探索)

EM4では、結果のウィンドウで、ユーザー側が「リフト(Lift)」、「信頼度(Confidence)」、「支持度(Support)」などの値を指定して、テーブルのサブセットを表示することができるようになりました。アソシエーションノードを適用した場合には、多数のルールが発見されることがありますが、この機能を用いると出力結果のうちに必要な部分だけを表示することができます。



[テーブルのサブセットの表示する範囲を指定する]

また、「全般」タブの「最小トランザクション度数/アソシエーションのサポート」欄において、もしくは、「連続 因果的連鎖」タブの「逐次サポートの最小トランザクション度数」欄において、パーセントで指定する項目が追加されました。

5.12 変数の選択ノード (探索)

データマイニングを行う際には、通常極めて多数の入力変数が存在します。このノードは、ターゲットとしている変数とあまり関連性がない変数を除外します。あらかじめ入力変数を減らすことにより、後の分析における処理時間を短縮させることができます。変数選択ノードで作られたロジスティックモデルに対して、分類を行うときの予測値のしきい値を設定できるようになりました。「ターゲット連関」タブで、「スコアデータセット」の欄を選択した場合に、「限界」の箇所に数値を指定してください。

6. その他

「変数」タブでの検索機能の追加

各ノードを開いて「変数」タブを選択したときに、例えば「測定水準」、「変数」などの上で右クリックをすることにより、検索を行うことができます。これは、変数が多数存在する場合、目的の変数を探すために大きな役割を果たします。

おわりに

以上、Enterprise Minerソフトウェア バージョン4.0における拡張点を中心にご紹介しました。上記以外にも、多くの拡張点がありますので、詳細につきましては、下記の新刊マニュアルをご参照ください。

Getting Started with Enterprise Miner Software, Version 4.0

注文番号：57436 (英語版)

価格：3,600円

Enterprise Minerソフトウェアは、データマイニングプロセスのすべての機能を統合して提供するソフトウェアです。本書は、Enterprise Minerソフトウェアで提供する基本的な機能を、初心者向けに説明したものです。各機能の概略と操作例を説明しています。これにより、各機能の操作方法と役割を習得することができます。

Data Mining Using Enterprise Miner Software: A Case Study Approach, First Edition

注文番号：57872 (英語版)

価格：2,800円

本書はEnterprise Minerソフトウェアのインターフェースを利用し、データアクセスから分析までを事例を用いて紹介した入門書です。Enterprise Minerソフトウェアの初心者にとって本書は非常に有益です。本書に記載されている事例を理解した後に、Enterprise Minerソフトウェアのオンラインリファレンスに記載されているより複雑な統計解析に取り組むと良いでしょう。

こちらに記載する価格はすべて税抜きです。ご購入については、専用のお申し込み用紙にてFAXで弊社「マニュアル販売係」までお申し込みください。

- TEL 03-3533-3835
- FAX 03-3533-3781
- E-mail booksale@jpn.sas.com



Q&A

複数のシートにまたがったデータを読み込む方法について(Windows版)
 DelphiからSASを起動する(Windows版)
 複数の外部ファイルをまとめてデータセットに取り込む
 配列を利用して、データ中の欠損値を探す
 UPLOAD、DOWNLOADプロシジャで数値変数の長さが増えるのを
 回避する(対象リリース: リリース6.08 ~ 6.12)
 大容量データを扱う場合の注意点(対象リリース: 8.1)
 SAS/CONNECTサーバーで起動するSASのバージョンとSPAWNERの
 組み合わせ(Windows版)
 SAS V8.1でプリンタの両面印刷に関するオプションを制御する方法
 (Windows版)
 SAS V8.1で編集中のプログラムの自動保存について(Windows版)
 OS/390 UNIX システムサービス(OE)移動時のSHAREサーバー用
 設定について(OS/390版)
 UNIX版SAS 標準ファイルディスクリプタの扱いについて
 (UNIX版)
 非線形混合モデルについて
 SAS/IMLソフトウェアでExcelのデータを直接利用したい(Windows版)
 SAS/IMLソフトウェアのV8elにおける拡張点について

Q Microsoft Excelの複数シートにまたがったデータを
 SASに読み込みたいのですが、どのようにすれば良い
 でしょうか？

A DATAステップを複数記述してマージすることもでき
 ますが、INFILEステートメントを複数記述すること
 より、一つのDATAステップで読込むことができます。

例)

```
filename test1 dde 'Excel|[Book1]Sheet1!R1C1:R3C2';
filename test2 dde 'Excel|[Book1]Sheet2!R1C1:R3C2';

data a;
  infile test1;
input a b;
  infile test2;
input c d;
run;
```

Q Delphiを利用してアプリケーション開発を行っています。
 その際、開発済みのSASプログラムを起動し、エンド
 ユーザーにはSASを意識させないようにしたいと考え
 ています。何か良い方法はありますか？

A DelphiのWinExec関数を利用して、バッチモードで
 SASを起動し、開発済みのSASプログラムを利用して
 処理を行ってはいかがでしょうか。下記にサンプルと
 してプログラムを作成いたしましたので参考にしてください。

サンプル例)

```
procedure TForm1.Button1Click(Sender: TObject);
begin
  WinExec('d:%sas%sas.exe d:%test.sas -nologo',SW_HIDE)
end;

end.
```

書式例)

```
WinExec('SASルート¥SAS.EXE SASプログラム -nologo',SW_HIDE)

-nologo : SAS起動時にロゴウィンドウを非表示にします。
SW_HIDE : ウィンドウを非表示にして処理を行います。
```

なお、ウィンドウの表示方法は他にも種類があります。ご利用用途
 に合わせてパラメータを変更してください。詳細に関しては、
 Delphiのオンラインヘルプ等をご参照ください。

Q 特定のディレクトリの下(サブディレクトリを含む)
 にまとまっている、同種のファイルを一つのデータセッ
 トとして取り込みたいのですが、どのようにすれば良
 いでしょうか？

A FILENAME ステートメントの PIPE エンジンを使って、
 ディレクトリ内のファイルの名前を取得し、その各々
 のファイルをINFILEステートメントのFILEVAR=オプ
 ションで指定することで、複数のファイルを一つのデータセットに
 取り込むことができます。以下のサンプルマクロは、指定したディ
 レクトリ以下にある、同一の拡張子のファイルをすべて取り込むも
 のです。

```
/* サンプルマクロ・プログラム */
filename cmd pipe 'dir /b /s c:%tmp%*.txt';

data result;
  length filepath$ 200;

  infile cmd;
  input filepath;
  infile dummy filevar=filepath end=done;

do until (done);
  input var1 var2 var3;
  output;
end;

run;
```

Q SASデータセット中の欠損データをもつオブザベーション番号と変数名の取得方法を教えてください。

データ例)

OBS	VAR1	VAR2	VAR3	VAR4
1	5	11	2	9
2	欠損	31	45	60
3	39	4	7	26
4	90	19	欠損	77

A 欠損値を持つオブザベーションの番号と変数名の取得方法について説明します。DATAステップを使用して、各オブザベーションの各変数の値をチェックします。ARRAYステートメントと、反復DOステートメントを使用すると、変数名の指定を簡略化できます。VNAMEルーチンを使用すると、配列要素となっている変数名を取り出すことができます。下記の例をご参照ください。この例では、データセットWORK.TESTの欠損値を持つオブザベーション番号と変数名をデータセットWORK.MISSINGに出力します。

```

/* WORK.MISSINGに欠損値の情報を出力します */
data work.missing;
  set work.test;

  /* DOループを使うために、配列を定義しておきます */
  array nvar _numeric_;
  array cvar _character_;

  length obs 8 vname $8;

  /* 数値変数の欠損値を調べます */
  do i=1 to dim(nvar);
    if nvar[i]=. then do;
      obs=_n_;
      call vname(nvar[i], vname);
      output;
    end;
  end;

  /* 文字変数の欠損値を調べます */
  do i=1 to dim(cvar);
    if cvar[i]=' ' then do;
      obs=_n_;
      call vname(nvar[i], vname);
      output;
    end;
  end;

  /* 必要な変数だけを出力します */
  keep obs vname;
run;

```

Q 8バイトより短い数値変数を含むデータセットをUPLOAD、DOWNLOADプロシジャで転送すると、長さが1桁増えてしまいます(たとえば4桁が5桁になります)。これを回避する方法はありますか？

A UPLOAD、DOWNLOADプロシジャでは、数値変数の精度の低下を防ぐために、8バイトより短い数値変数を自動的に1桁増やす仕様になっています。バージョン6のUPLOAD、DOWNLOADプロシジャには、数値変数の長さが増えるのを回避するオプションはありませんが、CIMPORTプロシジャで、「EXTENDSN=NO」を指定することにより、元の長さのままデータセットを移送できます。

例) データセットをアップロードする場合

ローカルホスト側でCPORTプロシジャを実行し、データセットを移送ファイルに出力します。

UPLOADプロシジャで、移送ファイルをバイナリモードでアップロードします。

リモートホスト側で「EXTENDSN=NO」オプションを指定してCIMPORTプロシジャを実行します。

```

proc cport data=_lib.data1 file='data1.xpt';
run;

rsubmit;
proc upload infile='data1.xpt' outfile='/xxx/xxxx/data1.xpt'
  binary;
run;

proc cimport data=_lib.data1 infile='/xxx/xxxx/data1.xpt'
  EXTENDSN=NO;
run;
endrsubmit;

```

Q Windows版SAS V8.1で大容量のデータを扱う場合の注意点を教えてください。

A SAS/CONNECTソフトウェアを使用して別のV8.1に対して、コンピュータサービスを用いて4GB以上のデータを作成する場合、データ作成を行う側のSASに次のオプションを設定する必要があります。

```
-SYNCHIO
```

このオプションを、接続時に用いるスクリプトファイル内で指定するか、環境設定ファイルで指定してください。

Q SAS/CONNECTソフトウェアを使用したクライアント・サーバー環境でSASを使用しています。このたび、サーバーにV8.1をインストールし、当面V6.12と共存した環境で使用していこうと考えています。何か注意する点はありますか？

A

以下の2点について確認をしてください。

1. スポーナーのバージョンについて

サーバー上でスポーナーが起動するSASのバージョンと、そのスポーナーのバージョンは同じである必要があります。

		スポーナーが起動するSAS	
		V6.12	V8.1
スポーナー	V6.12	OK	NG
	V8.1	NG	OK

V8.1のスポーナーあるいはV6.12のスポーナーだけで、両方のバージョンを使用することはできません。

2. SAS/CONNECTソフトウェアのスク립トファイルについて

スク립ト(通常"TCPWIN.SCR")はV6.12とV8.1を別にする必要があります。なお、スク립トファイル内でSASを起動するために「SAS」とだけ指定されているところは、フルパスで指定することをお勧めします。

例) SAS V6.12を起動する場合

以下は「D:\sas612\」にSAS V6.12がインストールされている場合を想定しています。

TCPWIN.SCRファイルをテキストエディタ等で下記のように変更してください。

変更する箇所はTCPWIN.SCRファイルの60行目付近になります。

```
(前) type 'sas -dmr -comamid tcp -device grlink
-noterminal -no$syntaxcheck' LF;
(後) type 'd:\sas612\sas -dmr -comamid tcp -device
grlink -noterminal -no$syntaxcheck' LF;
```

例) SAS V8.1を起動する場合

以下は「D:\Program Files\SAS Institute\sas\v8」にSAS V8.1がインストールされている場合を想定しています。

TCPWIN.SCRファイルをテキストエディタ等で下記のように変更してください。変更する箇所はTCPWIN.SCRの「ready:」ラベルの後、61行目付近になります。

```
(前) type 'sas -device grlink -no$syntaxcheck' LF;
(後) type "'d:\Program Files\SAS
Institute\sas\v8\nls\ja\sas" -config "d:\Program
Files\SAS Institute\sas\v8\nls\ja\sasv8.cfg";
type '-device grlink -no$syntaxcheck' LF;
```

(パスが「Program Files」のように空白を含む場合は、""で囲んでください。)

Q

プリンタとして「両面印刷」が可能な機種を使っています。SAS V6.12からV8.1に変更したところ、印刷の設定として存在している「両面印刷」のチェックボックスが有効になっています。V6.12ではこのチェックボックスを外せば次回起動時からは無効だったのですが、なぜでしょうか。

A

SAS V8.1よりプリンタに関するオプションとして「DUPLEX」システムオプションが追加されました。

V6.12までは「印刷の設定」を変更すれば「profile2」カタログに保存されるので、次回起動時からはそのまま使用することが可能でした。V8.1からはシステムオプションが優先されますので、「印刷の設定」を変更しただけではそのままご使用できない場合があります。

両面印刷に関するオプション

起動時から両面印刷を行う : DUPLEX オプション

起動時から両面印刷を行わない : NODUPLEX オプション(初期設定)

上記オプションを環境設定ファイル(SASV8.CFG)に指定する必要があります。

例) 両面印刷を指定する場合は以下の1行を環境設定ファイル(SASV8.CFG)に指定します。

```
-duplex
```

なお、印刷についてのオプションは以下のものがあります。

BINDING=	両面印刷時と同じしろの指定
BOTTOMMARGIN=	印刷の下余白の指定
COLLATE=	部単位で丁合印刷をするかどうかの指定
COLORPRINTING=	カラー印刷ができるプリンタの場合、カラー印刷をするかどうかの指定
COPIES=	印刷時のコピー部数
DUPLEX	両面印刷ができるプリンタの場合、両面印刷をするかどうかの指定
LEFTMARGIN=	印刷の左余白の指定
ORIENTATION=	印刷の向きを指定
PAPERDEST=	印刷の出力トレイを指定
PAPERSIZE=	印刷の用紙サイズを指定
PAPERSOURCE=	印刷の用紙トレイを指定
PAPERTYPE=	印刷の用紙タイプを指定
RIGHTMARGIN=	印刷の右余白の指定
TOPMARGIN=	印刷の上余白の指定

各オプションの設定に関しては、オンラインヘルプを参照してください。

Q

SAS V8.1でプログラムの編集作業を行っています。編集中に「自動保存を完了しました」のメッセージがステータス行に表示されますが、何を意味しているのでしょうか？

A

編集中のプログラムが自動保存されていることを意味しています。V8.1の拡張エディタ及びプログラムエディタには「自動保存」機能が追加されています。それぞれのエディタで以下のように動作が違います。

例) 拡張エディタの場合

保存先ディレクトリはWindowsのTEMPディレクトリになります。既存のプログラムを読み込んでいる場合は、「Autosave of プログラム名.\$AS」の名前で保存されます。例えば、test.sasというプログラムなら「Autosave of test.\$AS」の名前で保存されます。新規に作成したプログラムの場合は、「Autosave of Untitled.\$AS」の名前で保存されます。

例) プログラムエディタの場合

保存先ディレクトリは現在のSASのカレントディレクトリに保存されます。これは、ステータス行の右端に表示されるディレクトリです。デフォルトで保存される名前は「pgm.asv」になります。なお、自動保存の間隔は「10分間」となっていますが、次の方法で変更することができます。

1.メニューバーを使用する方法

[ツール] => [オプション] => [プリファレンス] でプリファレンスのBOXを表示させて、「編集」のタブを選択します。そこに表示される「自動保存するまでの時間」を任意の時間に変更します。

2.WAUTOSAVEコマンドを使用する方法

WAUTOSAVE INTERVAL=xx で「xx」の部分に任意の時間を設定します。「xx」の単位は「分」です。

Q IBM OS/390をV2.9にバージョンアップしたところ、SAS/SHAREソフトウェアのSHAREサーバーに以下のエラーが起き、TCP/IP接続ができなくなりました。何が原因でしょうか？

エラーメッセージ

```
ERROR:SASADMIN::: call failed,system error is ".
LSCX470 **** WARNING **** ERRNO = ESYS
Generated in SOCKET called from line nnnnn of
MAIN(VSOCKET), offset nnnnnn
Vendor-specific TCP/IP error condition (IBM TCP/
IP: errno=156).
```

A SASに含まれているSAS/C Transient libraryは、OS/390 UNIXサービス(OE:Open Edition)が有効となっているOS/390V2.4以上のリリースをご利用の場合、その利用ユーザーに対してOpen Editionのセグメントを必要とします。また、このエラーはOpen Editionでのファイルシステムに対する権限が無い場合にも出力されます。Open Edition側設定をご確認いただき、利用ユーザー(SHAREサーバーのJOB名等)用のセグメント定義をお願いいたします。これらの設定が正しく行われている場合、「SYS1.PARMLIB(BPXPRMxx)」(お客様の環境により異なる場合があります)に記述されている「MAXPROCSYS=」「MAXPROCUSER=」の設定値が適切でないことが考えられます。これらの設定値を増やしてみるをお願いいたします。

Q UNIX版SASシステムにおいて、シェルの標準入力や標準出力に対して入出力が可能でしょうか？

A UNIX版SASでは、標準入力、標準出力、標準エラー出力の各ファイルディスクリプタに対して、デフォルトで以下のようなファイル参照名を定義します。また、これらのファイル参照名はプログラム中で定義する必要はありません。

ファイルディスクリプタ	SASファイル参照名
標準入力	STDIN
標準出力	STDOUT
標準エラー出力	STDERR

これらファイル参照名を利用することで、リダイレクト機能などよりデータの入出力が可能になります。

例) ファイルの内容を行番号を付けて出力する

```
/* プログラム(sample.sas) */
data _null_;
file STDOUT;
infile STDIN dsd sharebuffers;
input _infile_;
put @1 _N_ z4. @6 _infile_;
run;

/* 結果 */
% sas sample.sas < source.txt
0001 * MAKE SAMPLE DATA
0002 data test;
0003 input a b c ;
0004 cards;
0005 1 2 3
0006 4 5 6
0007 ;
0008 run;
%
```

Q SASで非線形混合モデルを扱うことはできますか？

A バージョン8.1から、非線形一般化混合モデルを扱うためのNLMIXEDプロシジャがSAS/STATソフトウェアに追加されました。バージョン6においてサンプルマクロとして提供されていた%nlinmixed、%glimmixedマクロとは異なり、数値積分を伴う反復計算による周辺尤度の最大化に基づく推定もサポートされています。

Q SAS/IMLソフトウェアでMicrosoft Excelのデータを直接利用したいのですが、どのようにすれば良いのでしょうか？

A SAS/IMLソフトウェア上でも、DDEの機能を利用してMicrosoft Excelのデータを直接利用することができます。

DDEに関する詳細は、「Windows版 SASシステム使用の手引きバージョン6 Second Edition」P.118～に記載されておりますのでご参照ください。

以下にプログラム例を記載します。

```
FILENAMEステートメントを利用してDDEを指定しています。
Excelの9行3列目までをx 9行4列目をyに読み込んでいます。
xとyにデータを読み込んでいます。
```

SAS/IMLソフトウェアの各ステートメント(do data等)に関する詳細は、オンラインヘルプ、またはマニュアルをご参照ください。

・オンラインヘルプ

コマンドボックスにHELP IMLと記述し、実行すると、SAS/IMLソフトウェアに関するオンラインヘルプが起動します。そこで、Alphabetical Index of Commands and Functionsを選択すると、ステートメントに関するヘルプを参照できます。

・SAS/IMLソフトウェアのマニュアル

「SAS/IML Software Usage and Reference Version 6, First Edition」
(注文番号：56040 価格3,600円)

```
/** FILENAMEステートメントの指定**/
filename test dde
'Excel\C:\TEMP\Book1.xls\Sheet1!R1C1:R9C4';

/**IMLの実行**/
proc iml;
infile test missover;

/** データの読み込み**/
do data ;
input a b c y ;

xx=a||b||c;
x=x//xx;
y=y//yy;
end;
closefile test;

print x y;
quit;
```

Q SAS/IMLソフトウェアのV8eにおける新規追加の機能を教えてください。

A SAS V8eにおけるSAS/IMLソフトウェアの主な拡張点は、次の3つです。

- 1.ファイナンス向けの関数の追加
- 2.ロバスト回帰に関するサブルーチンの追加
- 3.多変量時系列解析に関するサブルーチンの追加

1については、フォワード・レートからスポット・レートを求めるSPOT関数、その逆の計算を行うFORWARD関数、また、最終利回り(yield to maturity)を求めるYIELD関数など、計7つの関数が追加されています。

2については、ロバスト回帰を行うLTSルーチン、その他にMCD (minimum covariance determinant) 推定値を算出するMCDルーチンが用意されました。

3については、VARMAモデルに対する対数尤度関数を計算するVARMALIKルーチン、VARMAモデルに従った時系列を生成するVARMASIMルーチン、ユーザーが設定した分散共分散行列の多次元正規分布に従う乱数列を発生させるVNORMALルーチンなど、5つのルーチンが新たに利用できるようになりました。

上記拡張点についての詳細は、「SAS/IML Software: Changes and Enhancements, Release 8.1」(注文番号：P58050 価格1,000円)をご参照ください。

SAS Training

V8e新機能のご紹介

「ODS初級」コース(Output Delivery System)

日時: 2/8(木)~2/9(金)2日間 10:00~17:00

料金: ¥58,000 (または、チケット)

前提知識: 「初級プログラミング」を受講済み、または同程度の知識のある方

受講対象: SASバージョン8eを使用して、説得力のあるビジネスレポートを作成したい方

内容: SASの新バージョン8eから提供されます、今一番ホットな機能ODS (Output Delivery System) の基本機能について解説します。このODSでは、プロシジャからの出力ページを制御したり、また、HTML形式へ出力できます。これにより必要な出力のみを取り出し、また出力すべてを電子化することで、Webでの配信を可能にしています。また、複数の出力を1ページにまとめるテンプレート機能や、テンプレートのカスタマイズを行うことで、より見やすい出力結果を作成することができます。

会場: 東京都中央区勝どき1-13-1イヌイビル・カチドキ7F
SASインスティテュートジャパン
東京本社トレーニングルーム

九州トレーニングコース開催のお知らせ

「マクロ言語入門」コースを開催

日時: 3/5(月)1日間 10:00~17:00

料金: ¥20,000 (または、チケット)

前提知識: 「初級プログラミング」を受講済み、または同程度の知識のある方

受講対象: SASでアプリケーションプログラムを作成される方

内容: SASでアプリケーションを作成する上で必要になるマクロ機能の基本的な使い方を解説します。このマクロ機能は、共通するプログラムのサブルーチン化や、データによって実行するDATAステップやプロシジャステップの制御、また、データによって変数名やタイトルなどを可変にできます。特に汎用性のあるプログラムを作成する上では、必須の機能です。

会場: 北九州市小倉北区浅野2-14-1 小倉興産KMMビル3F
SASインスティテュート ジャパン 九州営業所
(JR小倉駅北口より徒歩1分・西鉄バス小倉駅前下車徒歩3分)

お問い合わせ・お申込み先

〒104-0054 東京都中央区勝どき1-13-1イヌイビル・カチドキ7F
SASインスティテュート ジャパン トレーニング担当

● TEL 03-3533-3835

● FAX 03-3533-3781

● E-mail training@jpn.sas.com

New Publications

新刊マニュアルのご紹介

SASシステム バージョン8 入門ガイド

注文番号: 17551 (日本語版)

価格: 3,500円

本書を利用することで、単純で図解入りの、ステップバイステップ形式の指示に従うことにより基礎的なSASタスクについて学ぶことができます。データへのアクセスや管理、多様なレポートを使用するデータの表現、いくつかの統計方式を使ったデータ解析といった最新のSASシステム環境についても学ぶことができます。SASシステムをはじめ利用される方、およびSASインターフェースの新しい特徴について知りたい人々にとって本書は最適です。本書は「Getting Started with the SAS System, Version 8 (注文番号: 57551)」を翻訳したものです。なおPDF版は「SASシステム日本語オンラインヘルプ (CD-ROM)」の中に含まれています。

SASシステム バージョン8 新機能の紹介

注文番号: 17588 (日本語版)

価格: 6,500円

本書は、リリース6.12以降(メインフレーム版のOS/390、CMSのリリースは6.09E以降、OpenVMS VAXのリリースは6.09E以降)の各リリースからリリース8.0までのSASソフトウェアの変更点と拡張点について概要を説明しています。本書は「What's New in SAS Software for Version 8 (注文番号: 57588)」を翻訳したものです。原書では、バージョン7までの新機能とバージョン8の新機能を区別して記述してありますが、日本ではバージョン7がリリースされなかったため、すべてバージョン8eの新機能としてまとめてあります。この差分についての情報が必要な方は、原書を参照してください。なおPDF版は「SASシステム日本語オンラインヘルプ (CD-ROM)」の中に含まれています。

「SASシステム バージョン8 入門ガイド(注文番号: 17551)」および「SASシステム バージョン8 新機能の紹介(注文番号: 17588)」はセットでお求めいただくとお得です。(10,000円 → 8,500円) セットの場合は「SASシステム バージョン8 入門セット(注文番号: V8ST1)」とご記入のうえお求めください。こちらに記載する価格はすべて税抜きです。ご購入については、専用のお申し込み用紙にてFAXで弊社「マニュアル販売係」までお申し込みください。

● TEL 03-3533-3835

● FAX 03-3533-3781

● E-mail booksale@jpn.sas.com

なお、最新のマニュアル案内は弊社ホームページ(<http://www.sas.com/japan/manual/index.html>)にて公開しておりますのであわせてご利用ください。

SAS Technical News Winter 2001

発行
株式会社SASインスティテュートジャパン

テクニカルニュースに関するお問い合わせ先

テクニカルサポートグループ

TEL: 03-3533-3877

FAX: 03-3533-3781

E-mail: technews@jpn.sas.com



株式会社SASインスティテュートジャパン

www.sas.com/japan/

東京本社
〒104-0054
東京都中央区勝どき1-13-1
イヌイビル・カチドキ 8F
Tel 03 (3533) 6921
Fax 03 (3533) 6927

大阪支店
〒530-0004
大阪市北区堂島浜1-4-16
アクア堂島西館 12F
Tel 06 (6345) 5700
Fax 06 (6345) 5655

九州営業所
〒802-0001
北九州市小倉北区浅野2-14-1
小倉興産KMMビル 3F
Tel 093 (512) 5014
Fax 093 (512) 5016