November 1999

SAS Technical News

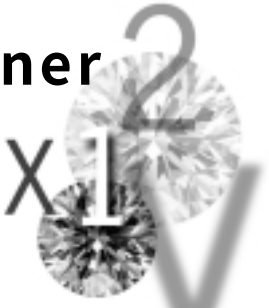
Volume 6 Number 2

Put the Power of the World's Leading Information Delivery System to Work in Your Organization.

CONTENTS

特集

Enterprise Miner ソフトウェア ～活用編～



データマイニングは、「大容量データから価値ある情報を発見するためのプロセス」と言うことができます。データマイニングでは、統計・機械学習・計算幾何などにおける手法が利用されており、その応用分野も金融・流通・マーケティング・品質管理・医療・通信など多岐にわたります。一般企業においては、多くの場合、蓄積されていた履歴データから、ビジネスにおいて有益な情報を導くために使われています。SASにおいてデータマイニングを包括的に行うプロダクトとしては、Enterprise Minerソフトウェアがあります。今までSASが統計分野で培ってきた技術が、Enterprise Minerには織り込まれています。ここでは、Enterprise Minerソフトウェアの特徴を簡単に紹介します。

1. データマイニングの手順

SAS Instituteではデータマイニングを、「データさえ入力してしまえば自動的に結果が出てくる」というものではなく、『一連のプロセス』と考えています。SAS Instituteでは、データマイニングにおける一連のプロセスを、"SEMMA"と言葉を使って表しています（"SEMMA"は、「セマ」と発音しています）。SEMMAとは、Sampling （データの抽出）・ Explore （データの探索）・ Modify （データの加工）, ・ Model （モデルのあてはめ）・ Assess （モデルの評価）の頭文字をとったものです。これら5つの処理を順に行なっていくことにより、より妥当なモデルをスムーズに作成することができるとSAS Instituteは考えています。この5つの処理について述べていきます。

1.1 Sampling－サンプリング

データマイニングでは、分析対象となるデータの大きさが、ギガやテラの単位になる場合があります。大容量データの全てを分析対象とすると、計算時間がかかったり、コンピュータのリソースが不足したりしてしまいます。そのような時には、「サンプリング（抽出）」を行なって、データの一部分だけを分析対象とすることが考えられます。扱いやすい大きさのデータにすることにより全データを解析するのに必要なコンピュータを用意せず、計算時間の短縮が実現できます。

1.2 Explore－探索

モデルを学習（推定）する前に、探索的な解析によって、データの大まかな傾向を予め把握しておくことは大切です。データの傾向を捉えるには、グラフによってデータの分布を眺めるなどの方法があります。また、分析対象のデータの変数が数百という大量の数になっている場合には、ターゲット変数と関連のあるものだけを残すという方法も考えられます。

1.3 Modify－加工

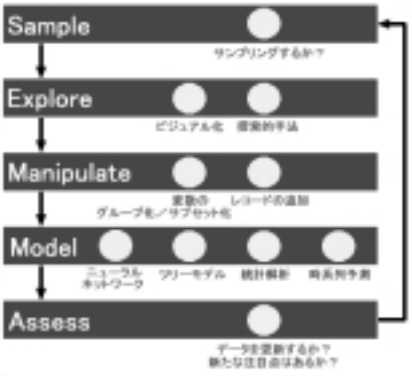
次に行う作業は、「データ加工」です。データ加工の例としては、分布の形状が歪んでいる変数を変換することなどが挙げられます。それ以外にも、複数の変数をまとめて一つの変数を作成したり、欠損値を何らかの値で補完することも考えられます。Sampling, Explore, Modifyという3つの作業が終了した時点で、モデル化を行なうための前準備が終了します。

1.4 Model－モデルのあてはめ

前準備が終わった後に、様々なモデルをあてはめます。モデルとしては、ニューラルネットワーク、決定木、回帰分析といったものがあります。各モデルには様々な設定方法があり、これらの設定を変更して、より適切なモデルをあてはめていきます。

1.5 Assess－評価

複数のモデルを推定したら、適切なモデルを選択する必要があります。いくつかの評価基準（例えば、利益の期待値・リフト率）に基づいて、実際のビジネスに促したモデルを探し出します。以上が、SEMMA プロセスの流れです。Enterprise Miner はSEMMAプロセスに従って設計、開発されており、分析者はこの流れに沿ってデータマイニングプロジェクトを進めることができます。



2. Enterprise Minerの特徴

ここでは、データマイニング統合パッケージであるEnterprise Minerがもつ７つの特徴を紹介します。

2.1 SEMMAモデルに従って開発されている

Enterprise MinerはSEMMAモデルに従って開発されているので、データマイニングの一連の流れをスムーズに実行することができます。

2.2 データソースを選ばない

データマイニングの分析対象となるデータは、さまざまなDBMS(データベースマネージメントシステム)に格納されているでしょう。Enterprise Minerは、SASのデータアクセス機能により、主要なDBMSにアクセスし、それらを透過的に利用することができます。

2.3 クライアント/サーバーで実行可能

Enterprise Minerはスタンドアローン環境でも実行可能ですが、大容量データの分析がスムーズに行なえるようクライアント／サーバー環境での実行もサポートしています。サーバー側にあるデータをサーバーのリソースを利用して処理を行ない、結果のみをクライアント側に表示させるといった事が可能です。

2.4 OS、ハードウェア環境を幅広く選択することが可能

Enterprise Minerは、さまざまなOSやハードウェア環境上で稼動します。サーバーとしてサポートしているのはHP-UX、Solaris、AIX、Windows NT Server（1999年10月15日現在）。クライアントはWindows NTおよびWindows 95です。

2.5 GUI環境で実行可能

Enterprise Minerソフトウェアは GUI環境を提供しています。SASシステムのプログラミングをご存知ない方でもポイント・アンド・クリックで、データマイニングを行なうことができます。GUI環境を通して実行された内容はそのままプログラムで保存することができますので、バッチ処理で実行することも可能です。また、分析した手順や結果はそのままGUIベースで保存できるので、ある人の分析過程を別の分析者が辿ることが容易です。

2.6 モデルが豊富

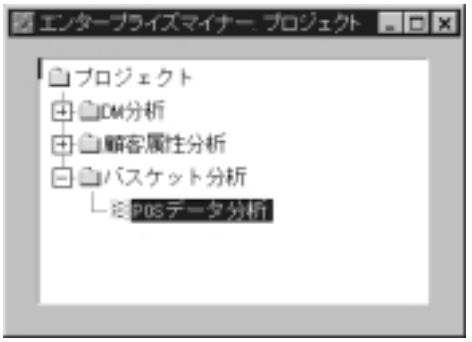
標準で、決定木や回帰分析、ニューラルネットワークといった手法が用意されています。また、通常のSASシステムで利用可能なプロシジャでプログラミングすることにより、ユーザ独自の新たなノードを作成することも可能です。

2.7 複数のモデルを同時に比較することができる

さまざまな手法を駆使して作成されたモデルは、最終的にどのモデルが最良なのかを評価する必要があります。Enterprise Minerには、作成された全てのモデルを１つのグラフに表現することによって、簡単に比較や検討することが可能です。

3. インタフェース

Enterprise Minerを起動すると、[プロジェクト]ウィンドウが立ちあがります。[プロジェクト]ウィンドウは、プロジェクトを階層形式で表示します。各プロジェクトには、ダイアグラムが格納されています。このダイアグラムがデータマイニングのひとつひとつのプロセスを保存しておくものです。



このダイアグラムを「プロセス・フロー・ダイアグラム （以下、PFD）」と呼んでいます。PFDが開かれると、[ツール]ウィンドウが立ちあがります。この[ツール]ウィンドウでは、PFDで実行可能な機能がノードとして表示されています。操作は非常に簡単です。これらのノードをドラッグ・アンド・ドロップし、ノードとノードを矢印で結んで、各ノードに自分が行ないたい分析の設定を指定するだけです。

4. ノードに含まれる機能

Enterprise Minerソフトウェア バージョン2.02では、様々なノードが用意されていますが、ここではそのうちの代表的なものについてご紹介します。

4.1 入力データソース

「入力データソース」ノードは、分析対象となるデータを指定するためのノードです。入力データソースノードを実行すると、分析対象のデータから標本が抽出されます。このデータのことを、Enterprise Minerでは“メタデータ”と呼んでいます。デフォルトでは、メタデータとして2,000標本だけが出力されます。次に示す処理は、このメタデータに基づいて行なわれます。

・1変量の記述統計量を計算する時
・棒グラフを描画する時
・変数における階層を決定する時（変数選択ノードにて）
・INSIGHTノード

入力データソースの「変数」タブでは、各変数の役割（ターゲット・入力・度数・IDなど）や測定水準（2値・間隔・名義・順序）を設定します。また、各変数に対して幾つかの記述統計量が計算されます。例えば、間隔変数であれば最大値・最小値・平均値・標準偏差・欠損値の割合・歪度・尖度が、名義変数や順序変数であれば水準数などの情報が表示されます。

4.2 サンプリング

「サンプリング」ノードは、データの抽出を行なうためのノードです。通常、データマイニングで使用されるデータの多くは、大容量の履歴データです。大容量データの全てを分析に用いると、非常に計算時間がかかります。一部の抽出されたデータだけを用いることによって、モデルの推定にかかる時間を大幅に短縮することができます。サンプリングノードには以下の5つの手法が用意されています。

・単純無作為抽出法
・系統抽出法
　系統抽出法とは、母集団から規則的に等間隔で抽出する方法です。第1番目、第(N+1)番目、第(2N+1)番目、、、というように抽出します。
・層別抽出法（層化抽出法）
　層別抽出法は、ある名義変数（例えば、性別）を層とみなし、その各層からある比率でサンプルを抽出するという方法です。この抽出方法では、各層の標本数における比を、母集団における比（例えば、男女の比）と同じにすることができます。
・最初のNオブザベーション
　先頭のオブザベーションだけを抽出します。
・集落抽出法
　集落抽出法とは、複数のクラスター（集落）から一部のものを選んで、その選択されたクラスターに属する全データを抽出する方法です。クラスターを抽出する際に用いる方法として、単純無作為抽出法、系統抽出法、最初のNオブザベーションを選択することができます。

4.3 データ分割

データ分割ノードでは、分析対象となるデータを2分割または3分割します。分割された各データは、モデルの評価を行なう目的で利用されます。データ分割ノードでは、データを次の3つに分割することが可能です。

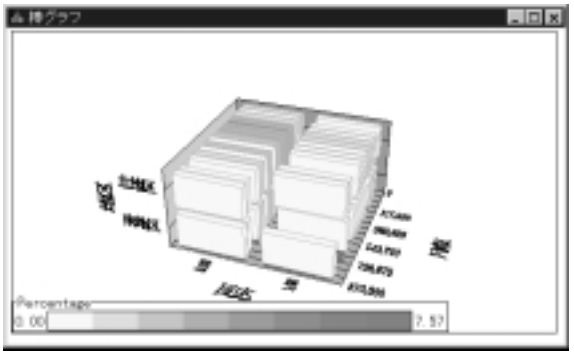
・学習用データ (training data)
　モデルを学習（推定）する時に使用するデータです。
・評価用データ（validation data）
　推定されたモデルの妥当性を確認するために使用します。モデルを学習および推定した場合、”over-fitting”(過度のあてはめ)もしくは”over-training”(過学習）と呼ばれる状況に陥る時があります。過度にあてはめられたモデルは、一般的なデータに対する推定精度が悪くなります。Enterprise Minerでは、過学習のモデルにならないようにするために評価用データが自動的に使用され、より一般性があるモデルが選択されます。
・テスト用データ (test data)
　各モデルの予測精度を計算したり、決定木・回帰・ニューラルネットワークといった異なったモデルを比較するために使われるデータです。

データ分割ノードでは、分割の割合や抽出方法を変更することができます。

・分割の割合
　元データを学習、評価、テストという３つのデータにどのような割合で分割するのかを指定します。デフォルトは、学習用データが40%、評価用データが30%、テスト用データが30%です。
・抽出方法
　次の抽出手法をサポートしています。
　◇単純無作為抽出法
　◇層別抽出法（層化抽出法）
　◇ユーザーが定義した抽出方法

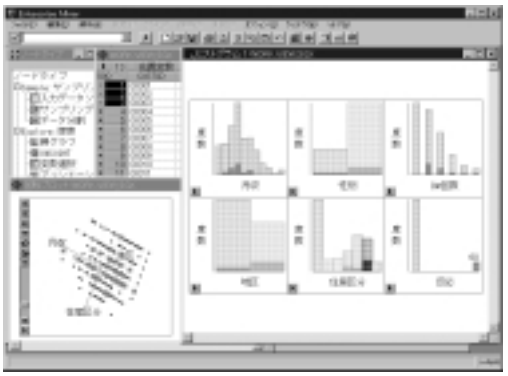
4.4 棒グラフ

棒グラフは、データを視覚化することによって、データの傾向を把握するのに使われます。Enterprise Minerでは最大3次元までのグラフを表示することができます。表示されたグラフはさまざまな角度に回転させることによって、平面では発見するのが難しい変数間の関係を明らかにします。変数のパターンやトレンドの発見、外れ値の発見に用いられます。



4.5 INSIGHT

SASシステムのビジュアルデータ解析ツールであるSAS/INSIGHTソフトウェアの機能です。SAS/INSIGHTソフトウェアは、ヒストグラムや散布図、主成分分析などを実行して、データの大まかな傾向を把握するのに役立ちます。



4.6 変数選択

データマイニングで使用する変数の数は数百、時には千単位になります。それら全ての変数を分析すると時間がかかってしまいます。変数選択ノードでは、ターゲット変数と関係がない変数を削除し、入力変数を減らすことができます。ここでの変数選択の基準は、（ターゲットが2値である場合も計算時間を短縮するために）線形回帰モデルのR2値（決定係数）に基づいて行われます。他にも、決定木の枠組みで X^2 値を選択基準とすることもできます。また、以下のような変数を分析対象から外すことが可能です。

・欠損値の割合が多い変数を削除する

　デフォルトでは欠損値が50%を超える変数を削除します。

・階層関係の変数を削除する

　非常に関係の深い項目（例えば 市と郵便番号）があり、情報が重複してしまうような場合に、両方とも分析に使用したくない場合があります。その際に階層構造を発見し、「詳細の情報」、または「最小の情報」のいずれかを保持するという指定を行ないます。

4.7 アソシエーション

複数の商品やアイテム間に存在する関連を調べるためのノードです。マーケティング分野において、「マーケットバスケット分析（買い物かご分析）」と呼ばれている分析を実行することができます。アソシエーションノードでは、次に示す２つの形式で、複数の商品における関連を調べます。

ⓐ「Aと同時にBが購入される確率はX%である」

ⓑ「Aを購入したお客様のなかで、Bも購入する確率はY%である」

マーケットバスケット分析においては、ⓐ のことを支持度(support)、ⓑ のことを信頼度（confidence）と呼んでいます。なお、分析を行う際に、「Aのすぐ後にBが購入される確率はX%である」というように時間（逐次性）を考慮することもできます。すべての組み合わせを考えると、膨大な数になることがあります。支持度や信頼度が低いものは出力しないようにし、できるだけ有益な情報だけを出力するように設定することができます。アソシエーションルールで計算された結果も、他のノード同様、データやHTML形式で保存できます。結果をそのままレポートにしたり、社内ホームページに載せることによって、情報を簡単に配信することが可能です。

4.8 データセット属性

データセット属性ノードでは、変数の属性を変更することができます。分析の途中で、分析で使っていなかった変数をターゲット変数に指定し直すことや、変数の測定水準を変更することができます。

4.9 変数変換

よりよいモデルを作成するには、モデル化の前に変数を適切な形で変換することが必要です。変数変換ノードでは以下の変数変換を行なうことができます。

・対数（log）
・平方根（sqrt）
・逆数（inverse）
・指数（exponential）
・標準化（standardize）
・ビン化 （binning）
　◇bucket（等間隔で分割）
　◇quantile（分位点による分割）
・ユーザーが定義した方法

4.10 外れ値

データマイニングを行なう場合にも、通常、予備解析によって「外れ値」を探します。他のデータとは傾向が違う「外れ値」は、モデルの学習を行う前に除外したほうがよい場合があります。外れ値ノードでは、ユーザーが指定した基準をもとに、自動的に外れ値を除外する機能があります。外れ値を除外する基準は以下の通りです。

・分類変数

　ある値の出現回数が指定した数以下のものは外れ値として除外します。

・間隔変数
・中央値からの平均絶対偏差

　◇中央値から離れている値を除外します。絶対平均偏差の何倍、離れているかを指定することができます（デフォルトは9倍）。

　◇最頻区間の中心からの偏差

　最頻区間の中心 (modal center) から、離れている値を除外します。

　◇平均からの標準偏差

　平均から離れている値を除外します。標準偏差の何倍離れているかを指定することができます。

　◇パーセント点

　上下のパーセント点がある値（デフォルトでは0.5%）以下のものを除外します。

4.11 データ置き換え

データマイニングで使用されるデータは、主に履歴データですから、良質なデータであるとは言えません。欠損値も数多く存在している

でしょう。回帰分析やニューラルネットワークといったモデル化のいくつかの手法では、1つでも欠損値を含むオブザベーションは除外して計算してしまいます。そのために、実際に使われるデータが非常に少なくなってしまう場合があります。データ置き換えノードでは、次の方法によって1変数毎に欠損値を埋めることができます。

・間隔尺度の変数
　◇平均値
　◇中央値
　◇範囲の中央（＝最小値＋範囲/2）
・名義尺度の変数
　◇最頻値

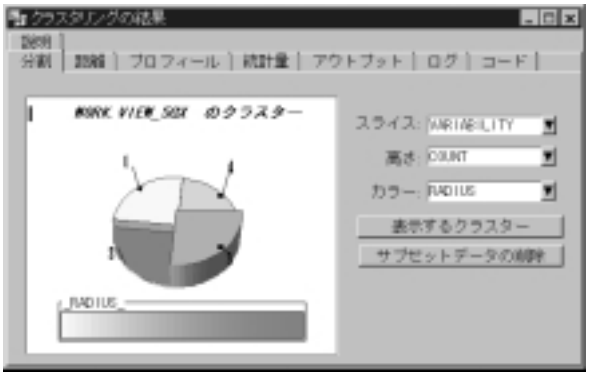
また、これ以外にもユーザーが指定した値で補完することも可能です。

「データ置き換えノード」では、欠損値以外の値を置換することもできます。例えば、東京、神奈川、埼玉、千葉を「関東地方」とし、大阪、兵庫、京都を「関西地方」というように置換することができます。ここで実行した処理によって元のデータベースが変更されることはありませんから、分析者が変数を分析しやすい値に自由に変更できます。

4.12 クラスタリング

クラスター分析は、似た属性をもつ標本が同じクラスターに属するように分類する手法です。クラスターに分類した後は、グループ処理ノードを利用することによって、個々のクラスター毎にモデルをあてはめることができます。クラスタリングノードでは実行する際に、例えば、以下の項目を選択することができます。

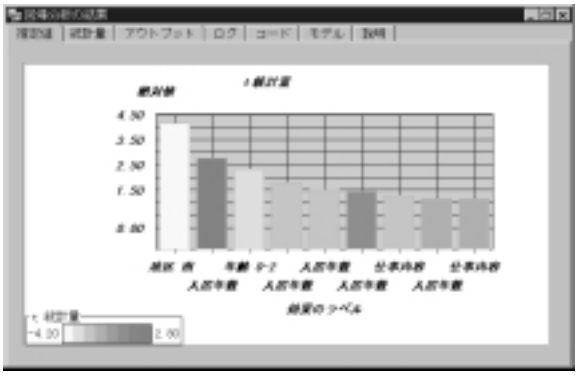
・クラスター数
・距離（ユークリッド距離、絶対距離など）
・欠損値の置き換え
・欠損値を含むオブザベーションを除外するかどうか



4.13 回帰分析

回帰分析は、統計解析でも頻繁に利用されてきた手法です。Enterprise Minerソフトウェアの回帰分析ノードでは、入力データソースノードにおいて設定されたターゲット変数や入力変数を自動的に判断し、線形回帰もしくはロジスティック回帰を実行します。回帰分析ノードでは様々な設定を指定することができます。例えば変数選択の方法としては、変数減少法（Backward）、変数増加法（Forward）、変数増減法（Stepwise）をサポートしています。また、変数選択の基準として、次のものをサポートしています。

・赤池の情報量規準
（AIC ; Akaike 's Information Criterion）
・Schwarzのベイジアン情報量規準
（SBC ; Schwarz 's Baysian informatin Criterion）
・評価用データにおける誤差（Validation Error）
・評価用データにおける誤分類率（Validation Misclassification）
・交差確認法に基づいて計算された誤差（Cross-Validation Error）
・交差確認法に基づいて計算された誤分類率
（Cross-Validation Misclassification）



4.14 決定木

データマイニングで頻繁に利用される手法として決定木があります。Enterprise Minerの決定木ノードでは分岐に使用する基準として次の3つを用意しています。

・X^2値のｐ値
　デフォルトでは、ｐ値が0.20以下の分岐までが探索されます。
・エントロピー
・ジニの多様性指標 (Gini 's diversity index)

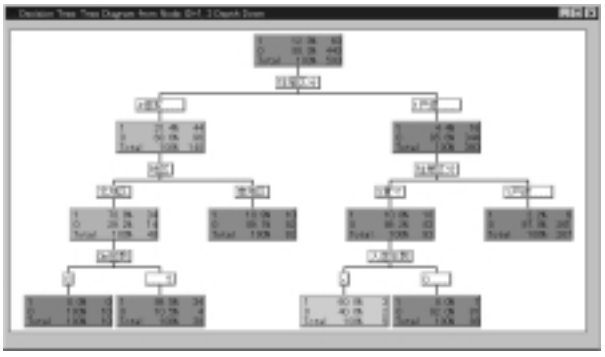
決定木を作成する時には、次の項目も設定することができます。

・葉に含めるオブザベーション数の最小値
・分割を行うオブザベーション数の最小値
・1つのノードから分岐される枝数の最大値
・決定木の深さの最大値
・代理変数の数

決定木では、自動的な学習（上記の3基準のいずれかに基づいて自動的に決定木を生成する方法）だけではなく、対話型の学習（分析者が分岐変数や分岐点を選択して決定木を作成する方法）もサポートしています。その他にも、決定木では以下の作業を行なうことも可能です。

・事前確率の指定
・利益もしくはコストを考慮した決定木の評価

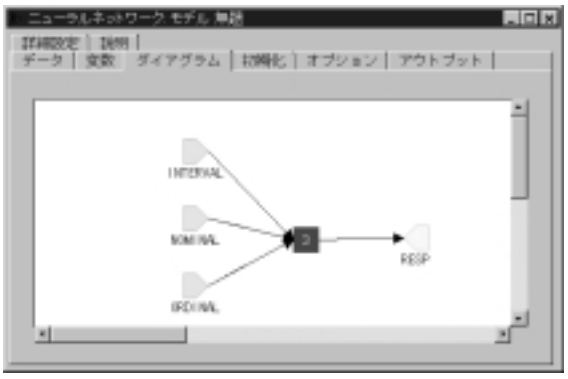
決定木の処理結果は、木のグラフ（どのような条件で分岐されているかを知るために使われる）だけではなく、リング形式のグラフ（データがどのような割合で分岐されているかを知ることができる）や、評価値を葉数ごとにプロットしたグラフ（決定木をどの程度の深さにするかを決めるために使われる）によっても示されます。



4.15 ニューラルネットワーク

ニューラルネットワークは、人間の神経生理学的な機能を模倣したモデルで、複雑な非線形の関係を表すのに適しています。ニューラルネットワークは非常に幅広いモデルを含みますが、Enterprise Minerでは、教師信号がある場合の階層的なニューラルネットワークをサポートしており、以下のようなモデルを指定することができます。

・一般化線形モデル (Genelized Linear Models)
・多層パーセプトロンモデル (MultiLayer Perceptrons ; MLP)
・動径基底関数モデル (Radial Basis Function; RBF)

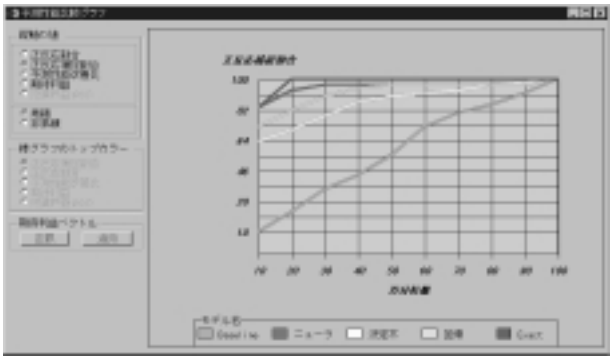


4.16 ユーザー定義モデル

Enterprise Minerでは、従来のSASシステムのプロシジャも利用したいというニーズに応えるため、ユーザー自身がプログラムしたモデルを使うこともできます。ユーザ定義モデルノードを用いることによって、SASのプロシジャをEnterprise Miner上で利用することができます。

4.17 アセスメント

アセスメントノードでは、利益の期待値・リフト率などのグラフに基づいて、様々なモデルの比較を行なうことができます。



4.18 スコア

スコアノードは、スコアリング（予測値の算出）を行うためのノードです。コードは、SASのプログラム（データステップ）で作成されます。スコアリングのためのプログラムには、単に推定されたモデル式だけでなく、それまでの事前処理（変数変換・データ置き換え・クラスタリング・グループ処理など）も含まれています。よって、新たなデータをスコアリングする時に事前処理を再び行う必要がありません。スコアノードで作成されたプログラムを実行するだけで、新たなデータに対してスコアを与えることができます。

5. 最後に

「ダイレクトメールの反応率が高い顧客を特定する」などのように、データマイニングを行なうには明確な目標を定めることが大切です。また、単にモデルを作成するだけではなく、反応率に違いはあったのか、反応してきた顧客は期待通りだったかなどを検討して、よりよいモデルを探索していく必要があります。データの取りこみや標準化などのクリーニングにかかる時間を短縮するためには、データウェアハウスの技術も組み合わせる必要があります。データマイニングは、データベースやマシンの環境や、ビジネス上の問題などを総合的に考えて行なう必要があります。Enterprise Minerはリリース以来、データマイニングの分野で、世界各国でさまざまなビジネス上の問題を解決しています。詳細については、弊社のホームページにて情報を提供しておりますので、是非、一度ご覧下さい。

●SAS Institute Inc.（米国SASインスティチュート）
http://www.sas.com/
●(株)SASインスティチュートジャパン
http://www.sas.com/japan/

Enterprise Minerソフトウェアに関するお問い合わせは下記までお願いいたします。

(株)SASインスティチュートジャパン
営業本部　TEL : 03-3533-6927
大阪支店　TEL : 06-6345-5700

SUGI-J '99
日本SASユーザー会総会 (SUGI-J '99)開催報告

1999年8月23日、24日の両日、東京全日空ホテルにて、SUGI-J '99が開催されました。両日合わせて1,042名のお客さまが来場され、すべてのイベントを盛会のうちに終えることができました。
SUGI-Jは、年に一度、SASユーザーの皆様にお集まりいただき、皆様の日頃の研究成果やビジネスにおけるSASの応用法などについての論文をご発表いただくとともに、SASインスティチュートジャパンから新機能や新バージョンのご紹介、SASのソリューションについての解説などを行なう一大イベントです。
本年度のSUGI-Jから、印象深かったものをいくつかご紹介します。

■論文発表

本年度は多くの分野にわたって42本の論文が発表され、その中から日本SASユーザー会世話人会の審査により、下記の各論文賞が選出、表彰されました。

●最優秀論文賞
「PROC GLM及びPROC IMLを用いた3期3剤クロスオーバーデザイン (直交ラテン方格) の解析」
ヘキスト･マリオン･ルセル株式会社 石川靖 氏
「混合正規分布によるVARモデル」
株式会社金融エンジニアリング･グループ 甲田恵氏、角谷督氏、加藤浩一氏
●世話人会特別賞
「尺度の最適変換を伴う回帰分析の適用事例」
専修大学商学部 町野正博氏、風間友太氏
●功績賞
東京大学医学部 浜田知久馬氏
日本ロシュ株式会社 高橋行雄氏
東邦大学医学部 田久浩志氏

功績賞とは、優秀な論文を数多く発表し、ユーザー間の情報交換に多大な寄与をした発表者を表彰するもので、今年度から設けられたものです。

■データマイニング特別講演

データマイニングの先駆者である米国SASインスティチュート ジョン･ブルックルバンクによるデータマイニング特別講演「Enterprise Miner Version 3.0の利用 ～ビジネス上の問題解決のために～」では、Enterprise Miner ソフトウェアのメジャーな機能拡張点について、データマイニングのフレームワークに基づいて評価されたいくつかの分析例を中心に、デモンストレーションを交えてご紹介しました。

■プレナリーセッション

これまで「日本SASユーザー会総会」として開催してきたものを、より発展させた形で開催されたものが「プレナリーセッション」です。プレナリーセッションは、ユーザー会初日の夕方、日本SASユーザー会 代表世話人、東京大学医学部 大橋靖雄教授、同じく副代表世話人、キリンビール株式会社 本川裕氏からのご挨拶と活動報告で始まりました。続いて弊社社長 デイヴィッド C. フェンダーからのご挨拶の後、理学博士 江崎 玲於奈氏 (ノーベル物理学賞受賞者、前筑波大学学長)より、特別講演「科学技術世紀の展望」として、人類が創生し発展させてきた科学技術についての興味深いご講演をいただきました。そして、論文賞の授賞式が執り行われ、各賞の受賞者に賞状と記念品が世話人会より手渡されました。なお、最優秀論文賞を受賞した2組には、副賞として2000年4月に米国インディアナポリスで開催される「SUGI 25」への招待券が弊社社長より手渡されました。

■フューチャーセッション

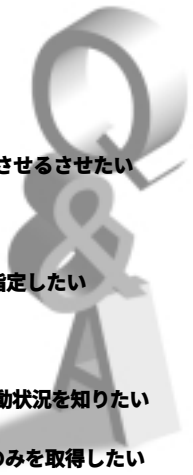
SAS Instituteが誇るSASシステムの次期メジャーバージョン「Nashville Project」。このコンセプトと広範な機能について、米国SASインスティチュート リサーチ＆デベロップメント副社長 Keith Collins、および同 アジアパシフィック リサーチ＆デベロップメント 董野真一郎よりご紹介しました。

■ハンズオン・ワークショップ

毎年 SUGI-J でご好評をいただいているハンズオン・ワークショップ。本年度は、75台という例年の倍以上のPCをご用意し、「SAS体験セミナー」「SAS Enterprise Minerセミナー」「データウェアハウスセミナー」「SAS/EISによるレポート作成セミナー」そして「SAS時系列予測セミナー」の5コースが開催されました。「SAS Enterprise Miner セミナー」では、今話題のデータマイニングツールを実際に操作できるということで、立ち見の方が出るほどの盛況でした。

SUGI-Jは、今回で18回目を迎えました。これから、おなじみのお客様にはもちろん、新しくお客様となられた方々にもご満足いただけるよう、お客様とSASインスティチュートジャパンとの接点として、有効に企画・開催していくよう努力していく所存です。

Q&A



●GPLOTプロシジャでID変数の値をグラフに反映させるさせたい
（SAS/GRAPH)
●任意のFONTをグラフ出力に反映させたい
（SAS/GRAPH）
●プロシジャで指定するデータセット名を可変で指定したい
（SAS基本機能）
●半角文字を全角に変換したい
（SAS基本機能）
●デュアルプロセッサーマシンでのSASシステムの稼動状況を知りたい
（SAS基本機能）
●SQLプロシジャを使用しDBMSの日付けデータのみを取得したい
（SAS基本機能）
●REGプロシジャでの変数選択による計算結果の違いについて
（SAS/STAT）
●MIXEDプロシジャのREPEATEDステートメントを指定した場合の注意点
（SAS/STAT）

Q GPLOTプロシジャで、プロットの横にIDとなる変数の値を出力したいのですが、可能でしょうか。

A ANNOTATE機能を使って、出力できます。

例
プロットの横に変数AGE（数値変数）の値を出力するANNOTATE機能の詳細は、「SAS/GRAPHソフトウェア リファレンス」マニュアルを参照してください。

```
                /* ANNOTATEデータセットの作成 */
data label(keep=x y xsys ysys text position style);
   set sasuser.class;          /* 入力データセット */
   xsys='2';                   /* x座標の単位系    */
   ysys='2';                   /* y座標の単位系    */
   position='6';               /* テキストの位置   */
   style='kanji';              /* フォント         */
   x=height+0.4;               /* x座標            */
   y=weight;                   /* y座標            */
   text=put(age,2.);           /* テキスト         */
run;                             ........

symbol1 v=dot c=blue;
axis1 label=(f=kanji);
proc gplot data=sasuser.class anno=label;

   plot weight*height /vaxis=axis1;
run;
```

Q SAS/GRAPHソフトウェアのフォント管理ユーティリティーを使って、WIN,WINPRTGドライバにWindowsのTrueTypeFontを登録し、'MSゴシック'などのフォントを使っています。PCによって文字タイプ番号が異なる場合がありますが、文字タイプ番号を合わせて登録できますか。

A フォント管理ユーティリティーを使用した場合、システムのフォント情報を自動的に取得するため、文字タイプ番号が異なる場合があります。その場合、いったんグラフィックドライバのエントリ（コピーされたもの）を削除して、使用するフォントだけを、番号を合わせて登録することをお勧めします。

GDEVICEプロシジャを起動して、文字タイプウィンドウで手入力で入力する方法もありますが、次のようなプログラムでも登録できます。

例
WIN,WINPRTGグラフィックドライバに、'MSゴシック'と'MS明朝'フォントを登録する

```
libname gdevice0 'd:\mydir';
proc gdevice c=gdevice0.devices nofs ;
   copy win       from=sashelp.devices;
   copy winprtg from=sashelp.devices;
      modify win
   charrec=(1,1,1,'MS ゴシック','Y')
   charrec=(2,1,1,'MS 明朝','Y');
      modify winprtg
   charrec=(1,1,1,'MS ゴシック','Y')
   charrec=(2,1,1,'MS 明朝','Y');
quit;
```

Q 任意のライブラリに保存されている全てのデータセットについてMEANSプロシジャを実行したいのですが、効率的な方法がありますか。

A Base SASソフトウェアのマクロ機能を使用すると良いと思われます。ライブラリの中のメンバー一覧は、SASHELP.VSTABLEビューで取得できます。下記の例は、次のようなことを行っています。

・SASHELP.VSTABLEから、任意のライブラリのデータを入力
・データセット名をDATAnにセット
・データセットの数をマクロ変数n_dataにセット
・データセット数がゼロのときは、データがないというメッセージを出力
・データセットがあるときは、データセット数分だけ、MEANSプロシジャを実行

なお、マクロに関する詳細は、「Base SASソフトウェアSASマクロ機能使用法およびリファレンス Version 6 Second Edition」を参照してください。

例

```
options mprint;
%macro all(lib);
  %let n_data=0;
  data _null_;
    set sashelp.vstable(where=(libname=%upcase("&lib")));
    n+1;
    call symput('data' || left(n),memname);
    call symput('n_data',n);
  run;
  %if &n_data = 0 %then
    %putデータがありません。もしくはライブラリが定義されていません。;
  %else %do;
    %do i=1 %to &n_data;
      proc means data=&lib..&&data&i;
      run;
    %end;
  %end;
%mend all;

%all(work)  ←――――― マクロの実行 ライブラリ参照名を指定
```

半角文字を全角文字に変換する関数などはありますか。

残念ながら、半角文字を全角文字に変換する関数はございません。しかし、半角文字を全角文字に変換するマクロを作成いたしましたのでご参考にしてください。

使用法

DATAステップの中で、呼び出してください。zentable, hantable, iの3個の変数を使用しています。

書式

```
%han2zen(元のテキストの変数名,変換後のテキストが入る変数名)
/*****************************************************/
/* han2zen: 半角テキストを全角に変換するマクロです。 */
/*****************************************************/
%macro han2zen(hantext,zentext) ;
  length zentable hantable $ 50 ;
  retain zentable 'ガギグゲゴザジズゼゾダヂヅデドバビブベボパピプペポ' ;
  retainhantable 'ｶﾞｷﾞｸﾞｹﾞｺﾞｻﾞｼﾞｽﾞｾﾞｿﾞﾀﾞﾁﾞﾂﾞﾃﾞﾄﾞﾊﾞﾋﾞﾌﾞﾍﾞﾎﾞﾊﾟﾋﾟﾌﾟﾍﾟﾎﾟ';

  &zentext=&hantext ;

  do i = 1 to length(hantable) by 2 ;
        &zentext= tranwrd
        (&zentext,substr(hantable,i,2),substr(zentable,i,2)) ;
  end ;

  drop zentable hantable i ;
  &zentext = ktranslate
  (&zentext,
  'ａｂｃｄｅｆｇｈｉｊｋｌｍｎｏｐｑｒｓｔｕｖｗｘｙｚ',
  'abcdefghijklmnopqrstuvwxyz',
  'ＡＢＣＤＥＦＧＨＩＪＫＬＭＮＯＰＱＲＳＴＵＶＷＸＹＺ',
  'ABCDEFGHIJKLMNOPQRSTUVWXYZ',
  '０１２３４５６７８９',
  '0123456789',
  'アイウエオカキクケコサシスセソタチツテトナニヌネノ',
  'ｱｲｳｴｵｶｷｸｹｺｻｼｽｾｿﾀﾁﾂﾃﾄﾅﾆﾇﾈﾉ',
  'ハヒフヘホマミムメモヤユヨワヲンァィゥェォャュョ',
  'ﾊﾋﾌﾍﾎﾏﾐﾑﾒﾓﾔﾕﾖﾜｦﾝｧｨｩｪｫｬｭｮ') ;
%mend han2zen ;

/************************/
/* han2zen の使用例です */
/************************/
  data test ;
    length a b $ 40 ;
      infile datalines ;
    input a &  ;
    %han2zen(a,b) ;
    put a= b= ;
datalines ;
SAS Institute Japan
ｻｽｲﾝｽﾃｨﾁｭｰﾄｼﾞｬﾊﾟﾝ
;
```

SASシステムリリース6.12を現在使用しています。デュアルプロセッサーのマシンを新たに購入することを考えていますが、どのくらい処理時間は短縮されますか。

SASシステムリリース 6.12 は、現在複数CPUをサポートしておりません。ただし、OSのシステムリソースをCALLしますので、OSが複数CPU対応であれば、全体的な処理速度は向上します。複数CPUの機能を使われる場合は、別途SPDServer 2.1 の導入をご検討下さい。このプロダクトは最新の並列処理機能とデータサーバ機能を使用しているため、多数のユーザを同時にサポートできます。

Q SQLプロシジャのパススルー機能を使用して　MS-ACCESSやORACLE等のテーブルの日付データを読み込むと、日時データになってしまいます。日付データとして読み込む方法はありますか。

A SASシステムのDATEPART関数を使用することで、日付値を取り出すことができます。以下の例では、ORACLEの日付データ hiredateをSAS日付値に変換します。

例

```
proc sql;
  connect to oracle (user=scott orapw=tiger path="@xxxxxx");
    create table data1 as
      select datepart(hiredate) as hiredate format=yymmdd8.
        from connection to oracle (select * from emp);
  disconnect from oracle;
quit;
```

Q REGプロシジャを用いて変数選択を行っています。変数選択を行った後の結果と、もう一度、REGプロシジャを用いて解析した結果とが異なっているのですが何が原因ですか。

A 変数選択の候補となっている変数に欠測値があるかどうかを確認して下さい。それらの変数において１つでも欠測値があるオブザベーションは、たとえ、最終的に選択されたモデルに含まれていないものであっても、計算から除外されます。そのため、再度、REGプロシジャによって、選択されたモデルをあてはめた結果と結果が異なってきます。

Q MIXEDプロシジャのREPEATEDステートメントを使って解析をしていると、次のようなメッセージが出力されて結果が出力されません。どうしてでしょうか。

```
An infinite likelihood is assumed in iteration 5
because of a nonpositive definite estimated R matrix for ID 1.
```

A このメッセージは、誤差の分散共分散行列 Rが反復計算の途中で非正定値行列になってしまい、尤度が無限大になったことを知らせるものです。R行列の対角要素が0もしくは負になると、御質問のようなメッセージが出力されます。特に、次のような状況において、メッセージが出力されます。

(a)入力データに間違いがある場合。入力データにおいて、1被験者内に同じ時点をもつオブザベーションが作成されている場合。
(b)推定するモデルに対して、データが相対的に少ない場合。

SAS Training

SASトレーニングのお知らせ

トレーニングサービスチケットに「5days」が仲間入り！

トレーニングを2コース以上受講予定の方に朗報です。お得なサービスチケットにお求めやすい5daysチケットが10/4より仲間入りいたしました。

価格及び有効期限：\950,000（受講日数5日分）
有効期限は使用開始日から6ヶ月間

本年度は、特別セミナー企画といたしまして、統計の基礎知識の習得を目的とした’初心者のためのデータ解析コース’、アプリケーション開発では欠かせないマクロ機能及びデータハンドリングを中心とした’マクロスペシャリストコース’、’実践データハンドリングコース’更に夜7:00から9:00までのイブニングスクールとして’データマイニング入門コース’を開催いたしました。各コースそれぞれお客様からの反響が非常に多くほぼ満席になりました。来期も同様にお客様からのご意見を参考に、トレーニングに反映していく予定ですのでご期待ください。

Latest Releases
最新リリース情報

■PCプラットフォーム

Windows版	SASシステムリリース 6.12 TS060
OS/2版	SASシステムリリース 6.12 TS020
Macintosh版	SASシステムリリース 6.12 TS040

■ミニコンピュータプラットフォーム

OpenVMS AXP版	SASシステムリリース 6.12 TS020
OpenVMS VAX版	SASシステムリリース 6.09E TS455

■UNIXプラットフォーム

MIPS ABI版	SASシステムリリース 6.11 TS040
Digital Unix版	SASシステムリリース 6.12 TS040
SunOSおよびSolaris版	SASシステムリリース 6.12 TS060
HP-UX版	SASシステムリリース 6.12 TS040
AIX版	SASシステムリリース 6.12 TS060
OpenVMS VAX版	SASシステムリリース 6.08 TS407

■メインフレームプラットフォーム

MVS版	SASシステムリリース 6.09E TS470
MSP版	SASシステムリリース 6.09E TS470
VOS3版	SASシステムリリース 6.09E TS470
CMS版	SASシステムリリース 6.08 TS410

New Publications
新刊マニュアルのご紹介

Learning SAS in the Computer Lab, Second Edition

●注文番号:57739（英語版）

●価　　格:4,500円

統計(コンピューター実習プロジェクトを含む)を学習している学生にとって、SASのシステムの基本を学ぶには、この革新的なマニュアルの第2版ぴったりです。著者は、実際のデータを分析し、重要な統計概念を教えるためにSASを使用することに焦点を当てて解説しています。最初の4章でSASで必要な事柄を学び、残りの18章が完全に独立しているので、任意の順に学ぶことができます。この第2版では、より多くの問題・データセット、およびロジスティック回帰、ノンパラメーター統計およびANOVAに関する情報を含んでいます。

Data Mining Techniques: For Marketing, Sales, and Customer Support

●注文番号:57699（英語版）

●価　　格:9,800円

データマイニング技術(Data Mining Techniques)は、データマイニングツールおよびデータマイニング技術の新しい世代を詳細に紹介し、よりよいビジネス上の決定を下すための利用方法を教えます。ビジネスデータのマイニング（採掘）の第1の実用的な手引きの1つは、マーケティング、販売およびカスタマー・サポート戦略を明確にするのに有用な顧客行動のパターンを見つけるための技術について記述しています。データベース分析者が彼らの好奇心を満たすために十分な技術情報以上のものを見つける一方、技術的に経験豊富なビジネスおよび販売責任者はその適用範囲を極めて入手しやすく感じるでしょう。以下のものに関するすべてを学習する機会があります。

- 北アメリカの主要な企業（leading company）は競争に打ち勝つためにどのようにデータマイニングを使用してるか。
- 各ツールがどのように働き、また仕事に適切なものをどのように取るか。
- 強力な7つの技術--クラスタ検知、メモリに基づいた推論、マーケットバスケット分析、遺伝的アルゴリズム、リンク分析、デシジョンツリーおよびニューラルネット。
- データマイニングのためにデータソースを準備する方法、および得られた結果を評価し使用する方法。

データマイニング技術は、休眠状態の情報システム内でビジネス解決のため金鉱を見つけ出す方法をすばやく容易に示します。

Year 2000
西暦2000年 年末年始 特別サポート体制のお知らせ

西暦2000年までいよいよあと1ヶ月あまりとなりました。弊社テクニカルサポートでは、Y2K問題に対応するため年末年始の1999年12月31日午後1:00から2000年1月5日の午前9:00まで、24時間の特別サポート体制を設置いたします。ご連絡方法につきましては、通常通りファクシミリ、電話、およびE-mailとなります。

●24時間特別サポート●

1999年12月31日 午後1:00

2000年 1月 5日 午前9:00

テクニカルサポートグループ
TEL : 03-3533-3877
FAX : 03-3533-3781
E-mail : support@jpn.sas.com

なお、期間中は、Y2K問題に関するご質問のみとさせていただきますので、ご了承いただきますようお願い申し上げます。

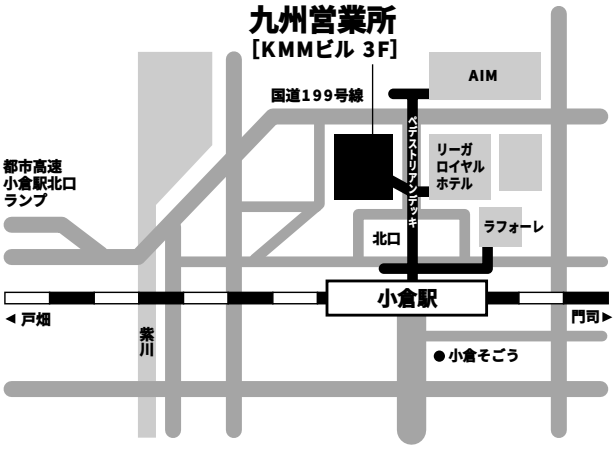
1999年12月28日(火)	通常営業 (～17: 00)
29日(水)	休業日
30日(木)	休業日
31日(金)	13: 00
2000年 1月 1日(土)	↑
2日(日)	
3日(月)	24時間特別サポート期間
4日(火)	
5日(水)	↓ 9: 00 (9: 00以降通常営業)

Information
九州営業所開設のお知らせ

SASインスティチュートジャパン、九州営業所を開設

去る10月1日（金）より、北九州市小倉にSASインスティチュートジャパンの九州営業所が開設されました。西日本へ向けたビジネスソリューションの展開だけでなく、地域企業に密接したきめ細かいサービスの提供をめざしていきます。九州営業所の所在地は次の通りです。

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



SAS Technical News
November 1999, Volume 6 Number 2

発行
株式会社SASインスティチュートジャパン

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: 03-3533-3877 FAX: 03-3533-3781

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

URL http://www.sas.com/japan/
NIFTY SERVE SAS Station:go sas