



**使用上の注意点 (Alert Notes)
Enterprise Miner (バージョン4.0)
Windows版**

はじめに

「使用上の注意点(Alert Notes)」は、インストールの前あるいはソフトウェアを使う上において、知っておく必要のある問題点を列挙しています。ご不明な点がございましたら、SASインスティテュートジャパン・テクニカルサポートまでご連絡ください。

1 導入時の注意点

1.1 Enterprise Minerソフトウェアのサーバとクライアントは、同一バージョン(Version4.0)である必要があります。Enterprise Minerソフトウェアは、異なるバージョン(Version2.0xとVersion4.0)によって、クライアント・サーバ環境を構築することはできません。

1.2 (Enterprise Minerソフトウェア Version2.01,2.02をお使いの方へ)Enterprise Minerソフトウェア Version4.0は、SAS System Version8.1と同じCD-ROMに含まれています。また、実行ファイル(sas.exe)も、SAS System Version8.1と同一ファイルになっています。実行ファイル(sas.exe)は、日本語版の場合、SASをインストールしたディレクトリの¥NLS¥JA¥配下にあります。

1.3 Enterprise Miner Version4.0をインストールするためには、[セットアップタイプ選択]ウィンドウにて、必ず「カスタム」を選択してください。「すべて」を選択しても、Enterprise Minerソフトウェアは選択されません。

[カスタムインストール]ウィンドウにて、次の3コンポーネントを必ず選択してください。

- a) 「SAS Enterprise Minerソフトウェア」
- b) 「SAS Enterprise Minerソフトウェア 日本語サポート」
- c) 「SAS Enterprise Minerソフトウェア DBCSサポート」

b)の「SAS Enterprise Minerソフトウェア 日本語サポート」は、「SAS 日本語サポート」で「オプション」プッシュボタンを押すことによって呼び出される「サブコンポーネントの選択」ウィンドウで選ぶことができます。また、c)の「SAS Enterprise Minerソフトウェア DBCSサポート」は、「SAS DBCSサポート」で「オプション」プッシュボタンを押すことによって呼び出される「サブコンポーネントの選択」ウィンドウで選ぶことができます。

2 Enterprise Minerソフトウェア ご利用時の注意点

2.1 アセスメントノードにおける問題

事前確率が定義されており、かつ、「データ分割」と「アンサンブル」ノードが使われている場合に、「アセスメント」ノードによって描かれる、検証(validation;評価)およびテスト(test)用のデータに対するグラフが間違っただものになる時があります。

メッセージなどによって、この問題が生じたことを知ることはできません。様々なモデルに対するグラフが正しい位置にプロットされておらず、また、ベースラインの直線も正しい位置にありません。なお、この問題が生じている場合も、「スコア」ノードによって作成されるコードには影響がありません。

この問題に関しては、SAS Note SN-003391に記述がございます。

現在、英語版の修正ファイルは提供されております。日本語版に関しては、弊社テクニカルサポートまでご連絡ください。

2.2 入力データソースノードにおける問題

SAS/ACCESSソフトウェアにおける機能のうち、LIBNAMEステートメントを用いて外部データベースへ接続している場合、それらのデータを「入力データソース」ノードによって読み込むことができません。この問題を回避するには、LIBNAMEステートメントではなく、SQLプロシジャのパススルー機能を用いるか、ACCESSプロシジャによってビューを作成してください。

この問題に関しては、SAS Note SN-003224に記述がございます。

現在、英語版の修正ファイルは提供されております。日本語版に関しては、弊社テクニカルサポートまで御連絡ください。

2.3 スコアリングにおける問題

SASシステムのオプションとして、VALIDVARNAME=V6を指定している場合、スコアリング用のコードを作成する処理が失敗します。Enterprise MinerではVersion8上だけで通用する変数名を作成するため、VALIDVARNAME=V6が設定されているとエラーが生じます。

また、Enterprise Miner Version4.0によって生成されたスコアリング用プログラムを、Version6で使用するためには、8文字以内の変数名に編集しなければいけません。

この問題に関しては、SAS Note SN-003255に記述がございます。

2.4 クラスタノードにおける問題

クラスタを行う基準(=距離の定義)として、「Midrange」、「Mean Absolute Deviation (Median)」、「Newton」、「Modified Ekblom-Newton」のいずれかを選択した場合、「クラスタ」ノードはスコアリング用のコードを正しく生成しません。そのため、これらの距離を選択した場合、クラスタに正しく各オブザベーションが分類されていません。また、「スコア」ノードにおけるコードも正しくありません。この処理が行なわれている時に、エラーやワーニング等は出力されません。

この問題を回避するには、「Least Squares」もしくは「Lease Squares (Fast)」をクラスタの基準として選択してください。

この問題に関しては、SAS Note SN-003218に記述がございます。

2.5 グラフィカルユーザインターフェイス全般における問題

デフォルトで設定されているフォントサイズを変更した場合(小さなフォントサイズに変更した場合)、タブウィンドウの内容が空白になる場合があります。また、Enterprise Minerが起動中に、フォントサイズを変更した場合、ウィンドウ内の配置が崩れて表示されます。

この問題に関しては、SAS Note SN-003256に記述がございます。

2.6 分析用データにおけるカテゴリ数の制限

カテゴリ変数の水準数が、intの最大値(2,147,483,647)以上存在する場合、そのintの最大値に丸められます。この制限は、データマイニング用のデータ(DMDB)を作成する時のものです。

この問題に関しては、SAS Note SN-003351に記述がございます。

2.7 レポーターノードにおける日本語処理に関する問題

[ツリー]ノードを用いた後、[レポーター]ノードにて、ツリーダイアグラム(樹形図)で表示する時に、日本語(2バイト文字)が文字化けすることがあります。その場合は、[ツリー]を再実行して、その後、[レポーター]ノードを再実行してください。

2.8 グラフィカルユーザインターフェイス全般における問題

メインウィンドウのウィンドウサイズを変更したのに、ウィンドウ内の各オブジェクトサイズが変更されない場合があります。ウィンドウサイズの変更がうまくいかなかった場合には、元のサイズへと直してください。

2.9 日本語訳の不統一について

“training(学習)”, “validation(検証)”, “test(テスト)”という3種類のデータのなかで、“validation data”の訳に対して、「評価」データおよび「検証」データという2種類の訳語が使われてしまっています。[ツリー],[回帰分析],[ニューラル]などのノードにおける[データ]タブの箇所では「評価」と表示されていますが、その他の箇所では「検証」と訳されています。これら2つは、同じ種類のデータを指します。

2.10 16バイト以上の日本語データに対する注意点

Enterprise Miner Version4の「回帰分析」・「ニューラル」・「ツリー」のノードでは、名義(nominal)・順序(ordinal)・2値(binary)に設定されたカテゴリ変数の文字を、初めの16バイトで識別します。17バイト目以降の文字は切り捨てられます。1バイト文字と2バイト文字が混ざった文字列が値としてあり、その文字列のちょうど16バイト目が、2バイト文字(日本語データ)であった場合、あふれた1バイトが切り捨てられ、有効な表示文字にならない場合があります。これが原因で、各ノードにおいて、スコアリング用のコードが正しく作成されなくなります。16バイト目には、2バイト文字を使わないようにしてください。

文字列の一部を、バイト数ではなく文字数によって取得する関数として、ksubstr関数が用意されています。この関数を用いることにより、初めの何文字かの文字列からなる変数をデータステップにて作成することができます。

2.11 ベータ版および評価版の機能について

次に挙げる機能は、ベータ版もしくは評価版です。また、日本語(2バイト)に対する処理を行いません。

C*Scoreノードは、英語版においてもベータ版です。このノードは、C言語のソースコードを生成するものですが、生成されたCのコードには、2バイト文字の文字列処理を考慮していない部分があります。このノードは、Enterprise Minerプロダクトとは、別プロダクトです。

ツリービューアは、Enterprise Miner Version4.0においては評価版です。これは、SAS/AFプロダクトにおける組織図オブジェクトではなく、Microsoft Foundation Classのオブジェクトを用いて開発されたビューアです。

ツリーの結果(ツリーダイアグラム)を見るために用います。このビューアは、日本語などの2バイト文字があると文字化けを起こします。

2.12 英語オンラインマニュアルにおける数式の間違えについて

英語マニュアルにおいて、「回帰(regression)」、「ツリー(tree)」、「ニューラルネットワーク(neural network)」、「予測モデル(predictive modeling)」の章において、数式が間違えております。多くの数式がずれて配置されています。日本語オンラインヘルプにおいて、修正を行なう予定です。

2.13 ターゲット変数に依存したアセスメントの問題

ターゲット変数が分類変数で、その変数名の長さデータ値の長さを足して14バイトを越える時、[アセスメント]ノードにおける結果が正しくないものになる可能性があります。この問題が生じたことは、混同行列などの行数や列数が、ターゲット値の水準数と異なることにより判断することができます。問題を回避するには、ターゲット変数の変数名の長さ、およびそこに含まれている文字値の長さを短くして、足して14バイトを越えないようにしてください。

この問題に関しては、SAS Note SN-003703に記述がございます。

2.14 [外れ値のフィルター]における問題

[外れ値のフィルター]ノードにおいて、[分類変数]や[間隔変数]タブにて、[すべての行を選択]もしくは複数行を同時に設定した場合、間違ったデータが作成されることがあります。

Enterprise Minerで表示されるテーブルでは、マウスでドラッグするか、[編集(E)]-[すべての行を選択(A)]で、複数行を選択することができます。この操作によって、[分類変数]タブで[除外する値の設定]、もしくは、[間隔変数]タブで[取り入れる範囲の設定]を選んだ場合、[外れ値のフィルター]ノードは正しくオブザベーションを除外しない時があります。除外が正しく行われていない場合にも、ノード自体では正しく処理が行われたという旨のメッセージや結果を出します。そのため、除外が正しく行われた結果かどうかを確認するための唯一の方法は、実際に除外がうまく行われたかどうかを、後からチェックするしかありません。問題を回避するには、1度に1列しか設定しないようにしてください。

例えば、データセット中に、X1とX2という2つの分類変数が存在しており、それぞれが'A'という値を含んでいたとします。[分類変数]タブにおいて、マウスで行をドラッグするか、もしくは、[編集(E)]-[すべての行を選択(A)]によって複数行を選択した後、[除外する値の設定]で表示されるヒストグラムで'A'を選んだとします。この時、[分類変数]タブのウィンドウ上では'A'という値が除外されることを示していますが、実際には2変数のうちのいずれか1つしか除外の条件に使われません。そのため、作成されたデータは間違ったものになります。

この問題に関しては、SAS Note SN-003580に記述がございます。

2.15 [データの分割]で層別抽出を行う時の問題

[データの分割]ノードにて、[層別抽出法]を選択した場合、ユーザが設定した割合とは異なった値で、学習・検証・テストのデータ分割を行う可能性があります。変数のある値の割合が非常に小さな場合、その違いが大きくなります。例えば、100,000レコードから10レコードを抽出する場合、7:3:1の割合で学習・検証・テストを

分けようとしても、この割合で分割されない場合があります。どのような割合になるかは、乱数に依存しています。この問題が生じたかどうかを知ることは、ノードを実行した後にも知ることはできません。エラーなどのメッセージを出力しないため、分析が終わった後に、各層の度数を算出して、問題が生じたことを初めて確認することができます。

この問題に関しては、SAS Note SN-003570に記述がございます。

2.16 [アソシエーション]ノードで時間順序を考慮した分析を行った時の問題

[アソシエーション]ノードで逐次分析を行う場合、[トランザクションウィンドウの長さ]で間隔を指定したとき、幾つかの関係が検出できていない可能性があります。

[アソシエーション]ノードで逐次(sequence)を選択した時、デフォルトでは[トランザクションウィンドウの長さ]は[最大間隔]になっています。もし、[使用する間隔を指定する]にて値を指定した場合、幾つかの関係が検出できていない可能性があります。この時、すべての統計量は間違っって出力されますが、明らかに間違いが生じているかどうかを知る方法はありません。

この問題が生じる条件を例によって説明します。ある顧客が、商品A, B, C, D, Eを次の順序で購入したとします。

<u>Sequence</u>	<u>Target</u>
1	A
2	B
3	C
4	D
5	A
6	E

ターゲット変数に、商品Aが2回購入されていることに注目してください。もし、[トランザクションウィンドウの長さ]が、同じ商品を購入するまでの長さ以下の場合、2回目に登場した商品の関係を検出できません。この場合だと、4以下の値を指定すると、A=>Eという関係が検出されません。この不具合が生じているかどうかを知るためには、1つ1つの顧客についての情報をチェックする必要があります。次の条件を満たしている場合、このバグが生じます。

- 1) 1顧客あたりのターゲット変数において、同じ値が2度以上、存在する場合
- 2) 複数個、存在するターゲット値の間隔以下の値を、[使用する間隔を指定する]で設定した場合

この問題に関しては、SAS Note SN-003568に記述がございます。

2.17 [データ置き換え]ノードにおける問題

分類変数が数値型で欠損値をもつ場合に、[データ置き換え]ノードにおいて[値の指定...]を行なうと、間違えた置き換えが行われます。

[データ置き換え]ノードの[分類変数]タブに、[欠損値を帰属させる方法](Imputation Method)という列があります。この列のある行で右クリックし、[手法の選択...]-[Set Value...]を選ぶと「Specify Value」ウィンドウが呼び出されます。このウィンドウの[データの値]の下矢印を押すと、データセット中の値がリスト表示されます。そこで値を選択し、[OK]ボタンを押すと、選択したデータ値が[欠損値を帰属させる方法]の列に表示されます。

しかし、この処理を実行しても、選択したものと違う値で置き換えが行われる可能性があります(実際には、前に一度選択した値が置き換えられてしまう)。この問題は、数値型の分類変数で、選択したもののなかに欠損値が含まれている時に生じます。

問題が生じていても、そのことを示すエラーやワーニングは出力されません。問題を回避するには、データセット中の値をリストから選択するのではなく、[その他の値]で値を手入力してください。

この問題に関しては、SAS Note SN-003669に記述がございます。

2.18[変数変換]ノードにおける問題

一度定義された入力データソースを変更しても、[変数変換]ノードが、その変更を反映しないことがあります。

次のようなフローを作成したとします。

入力データソース → 変数変換 → 回帰・ニューラル・ツリーなどのモデル作成用ノード

また、[入力データソース]で定義したデータセットに、変数Xが存在したとします。そして、一度、[変数変換]ノードで何らかの変換をした後、[入力データソース]ノードに戻り、変数Xを“reject”(除外)したとします。この時、[変数変換]ノードを実行すると、Xを変換した変数を除外するべきなのですが、除外されません。そのため、モデル作成用ノードで作ったモデルは正しくなくなります。また、後続の[アセスメント]や[スコア]の結果も間違ったものになります。この問題が生じた時もエラーやワーニングは出力されません。モデル作成用の各ノードで除外した変数が含まれていないかどうかをチェックすることによって、この問題が生じていないことを確認できます。モデル作成用の各ノードで、入力変数が正しく定義されているかを確認してください。

この問題に関しては、SAS Note SN-003489に記述がございます。

2.19[外れ値のフィルター]ノードにおける問題

[外れ値のフィルター]ノードを一度実行した後、入力データソースの設定を変更し、[外れ値のフィルター]を再実行すると、間違った条件によって除外が行われます。この問題が生じてても、エラーやワーニングは出力されません。問題を回避するには、[入力データソース]の定義を変更した場合には、[外れ値のフィルター]を一度、削除した後、ノードを新規作成してください。

この問題に関しては、SAS Note SN-003631に記述がございます。

3 その他

SASインスティテュートジャパンは、製品出荷後に発見された不具合に対する修正モジュールやマニュアルに記載されていない最新の情報等を次のサービスにより提供しています。

3.1 @nifty SAS Station

GO SASコマンドでSASステーションへアクセスし、

[4.データライブラリ] → [メンテナンスモジュール]を選択して最新の修正モジュールを入手することができます。

また、[4.データライブラリ] → [ドキュメント]を選択して最新情報等をご覧頂けます。

3.2 WWW(World Wide Web)

WWWブラウザを用いて、SASインスティテュートジャパンのホームページ (<http://www.sas.com/japan/>) にアクセスし、

[テクニカルサポート] → [FTPサービス] → [FTPサービス---サーバに接続] → [maint/] を選択して最新の修正モジュールを入手することができます。

また、[テクニカルサポート] → [SAS FAQ]、[使用上の注意点]を選択して最新情報等をご覧頂けます。

3.3 Anonymous FTPサービス

ftpコマンドを用いて、米国 SAS Institute Inc. の FTP サーバ ([ftp.sas.com](ftp://ftp.sas.com)) にアクセスし、
/pub/webfiles/Japan/maint ディレクトリに移動して、最新の修正モジュールを入手することができます。

ユーザID、パスワードは以下のように入力してください。

```
% ftp ftp.sas.com
ftp> Name(ftp.sas.com:userid): anonymous
ftp> Password: お客様のE-mailアドレス
ftp> cd /pub/webfiles/Japan/maint
```

上記サービスをご利用いただけないお客様で、修正モジュールが必要な場合には、弊社テクニカルサポートまでお問い合わせください。

Enterprise Miner(バージョン4.0)使用上の注意点(Alert Notes)

2000年11月21日 第2版第2刷発行

発行元 株式会社SASインスティテュートジャパン

〒104-0054 東京都中央区勝どき1-13-1 イヌイビル・カチドキ8F

電話 03(3533)3760

本書の内容に関する技術的なお問い合わせは下記までお願い致します。

SASテクニカルサポート

TEL:03 (3533) 3877

FAX:03 (3533) 3781