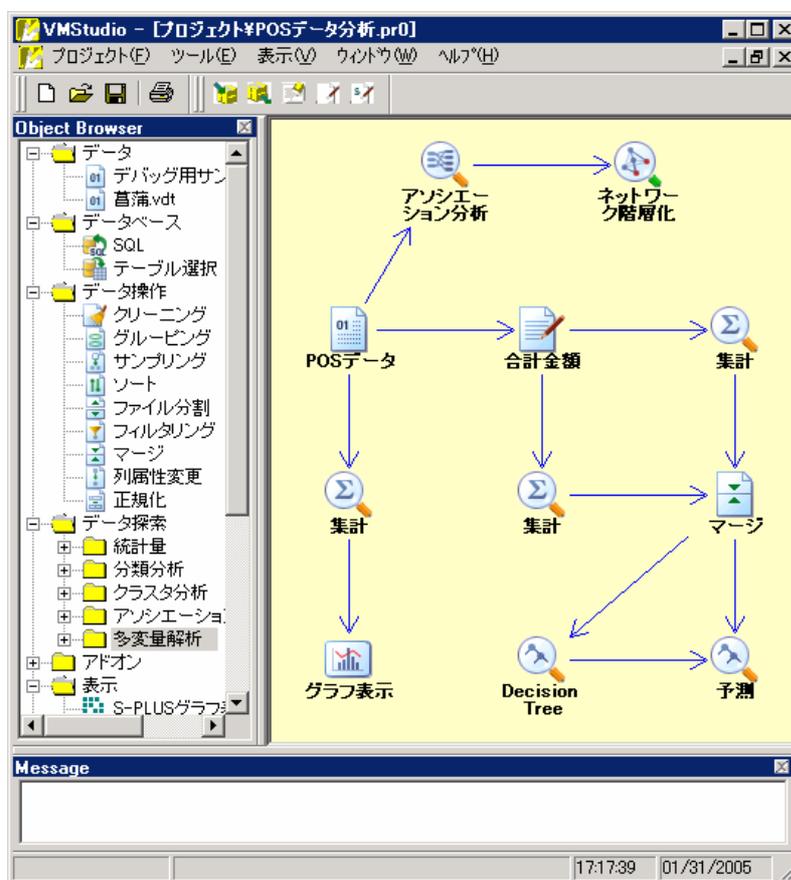


汎用データマイニングツール

Visual Mining Studio[®]

Version 5.0 新機能紹介



Ver.5.0 新機能一覧

インターフェース刷新

- 実行過程表示
- Undo 機能
- 自動保存
- 計算時間測定
- 外部エディタ呼出

多ユーザ対応の共有機能

- オンラインに対応

データベース更新機能

- 結果の反映

連携機能の追加

- S-PLUS 6
- R
- SAS

決定木の改良

- 予測時の欠損値対応
- 外部指定の強制分岐

Rule Base システム

- 手動で修正・追加

統計解析機能

- 主成分分析
- カーネル主成分分析

階層型クラスタリング

スクリプト関数大幅追加

1. ユーザーインターフェースの大幅刷新

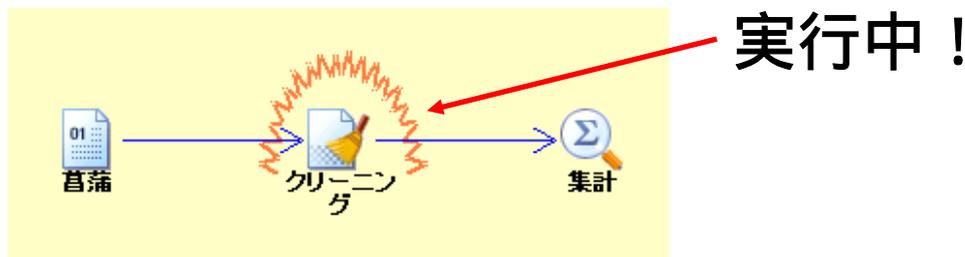
(ア) アイコンのデザインの刷新

性能を視覚的に表現したアイコンに変更したため、機能の視認性が大幅に向上いたしました。



(イ) 実行中のアイコンの強調

現在処理を実行しているアイコンを視覚的に認識できるようになりました。このため、長い処理フローの中での進行状況を視覚的に確認できるようになりました。

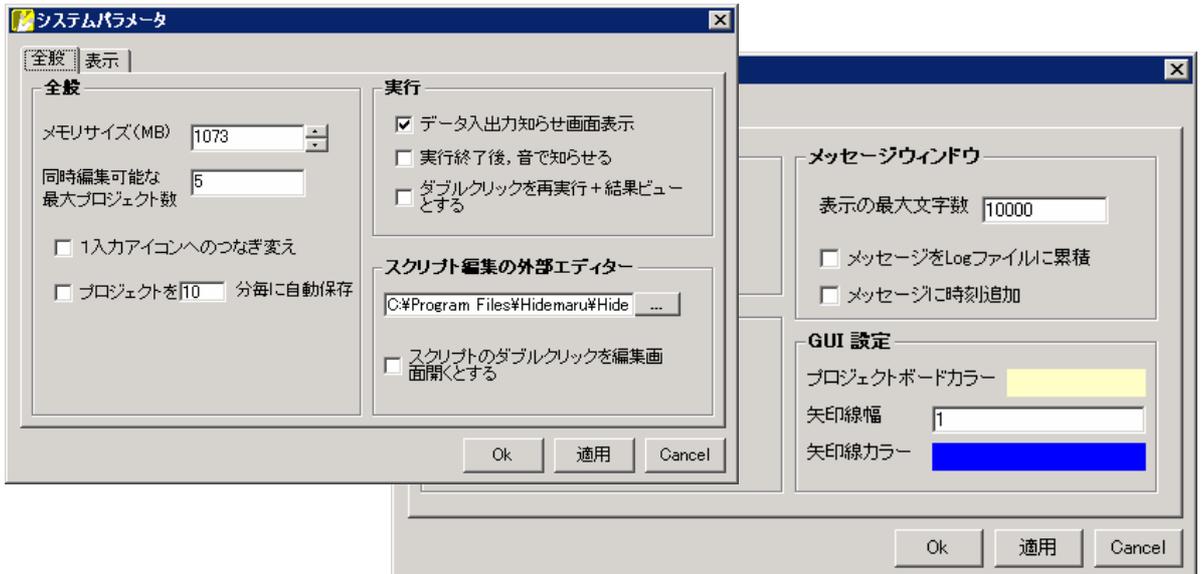


(ウ) Undo 機能

プロジェクトを作成中に誤った操作をしてしまったときなど、簡単に元の状態に復元できます。

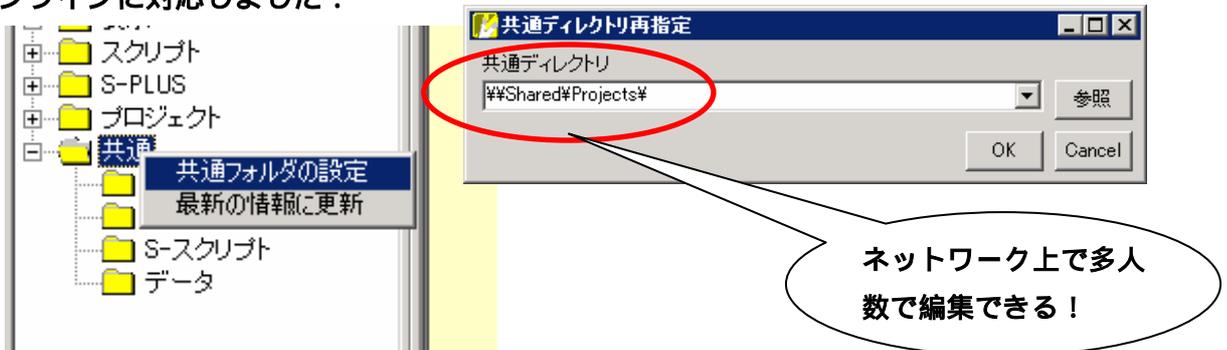
(エ) システム指定オプションの充実

ユーザ自身の好みに合わせて、環境を細かく設定できるようになりました。特に要望が多かった、プロジェクトの自動保存機能や計算時間計測機能・外部エディタの呼び出し機能が追加されました。



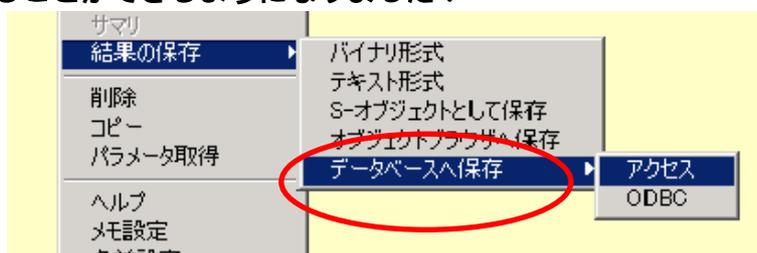
2. 他ユーザ対応のデータ共有機能

多人数で分析を行う場合には、通常共有ディレクトリを利用します。この機能がオンラインに対応しました。



3. データベース更新機能

データベースを更新することが可能になり、分析結果をデータベースに保存することができるようになりました。



4. 連携機能の強化

(ア) S-PLUS 6.1 J

汎用データ解析ソフト S-PLUS の日本語版最新バージョンである 6.1J と連携できるようになりました。S- オブジェクトとしてのデータの受け渡しについても、これまでと同様シームレスに行うことができます。

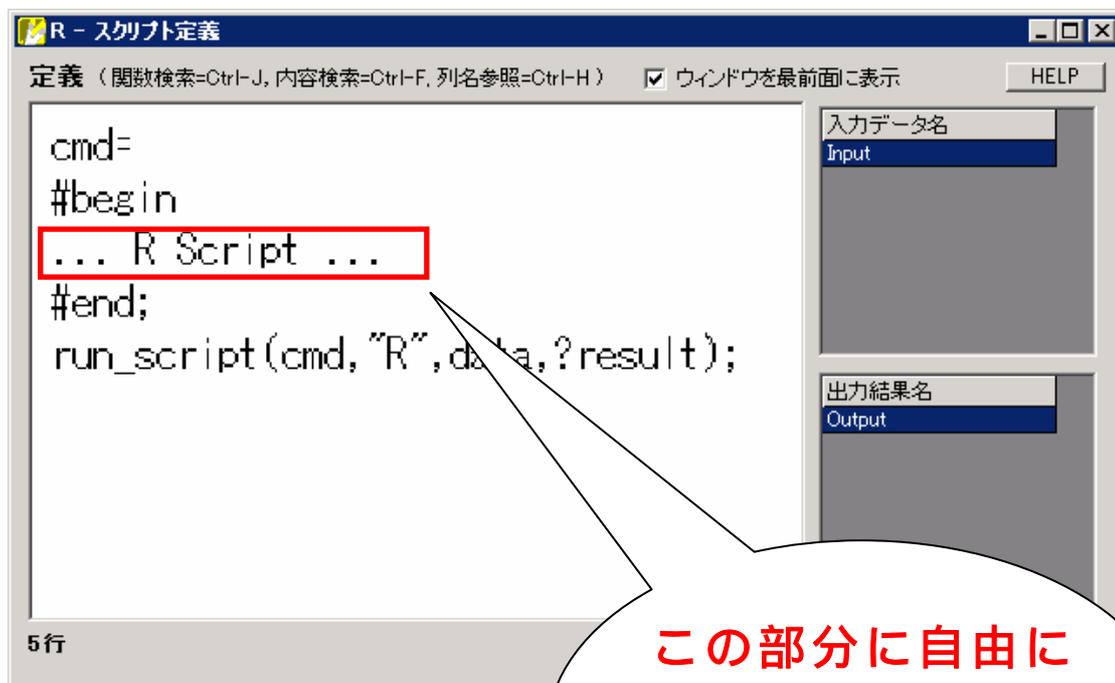
(イ) R (バージョン 2.0)

フリーのデータ解析ソフト R (バージョン 2.0) の機能呼び出すことが可能になりました。(注: R があらかじめインストールされている必要があります)

(ウ) SAS

米 SAS 社のデータ解析ソフト SAS の機能呼び出すことが可能になりました。(注: SAS があらかじめインストールされている必要があります)

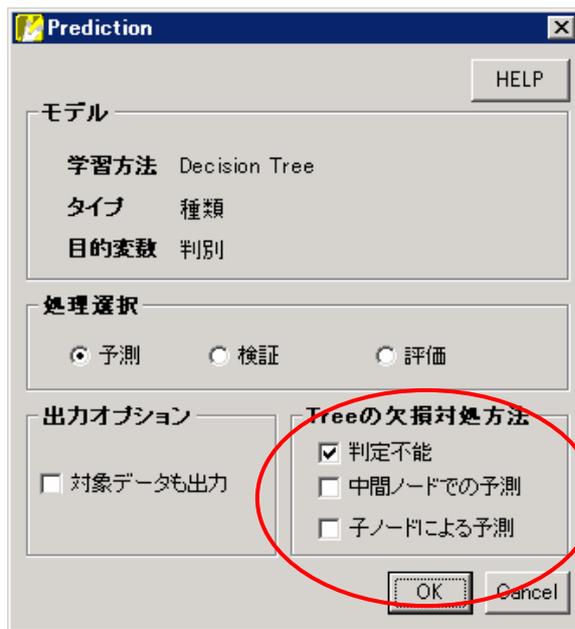
(例) VMStudio スクリプトによる R の呼び出し方法



5. 決定木の改良

(ア) 予測時の欠損値対応

予測時における欠損値の問題に対し、判定不能とするか、中間ノードでの予測を行うか、子ノードによる予測を行うか、の3通りの対処法が可能になりました。



判定不能

分岐不能の場合はあきらめる

中間ノードでの予測

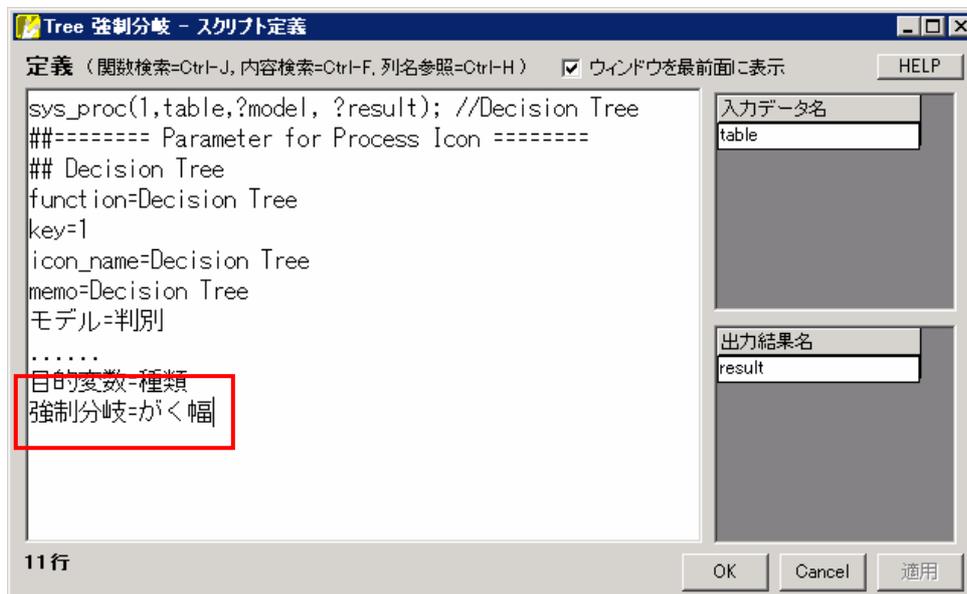
分岐不能の中間ノードのクラスを予測値とする

子ノードによる予測

分岐不能の中間ノードから重みを按分して、予測を先に進める

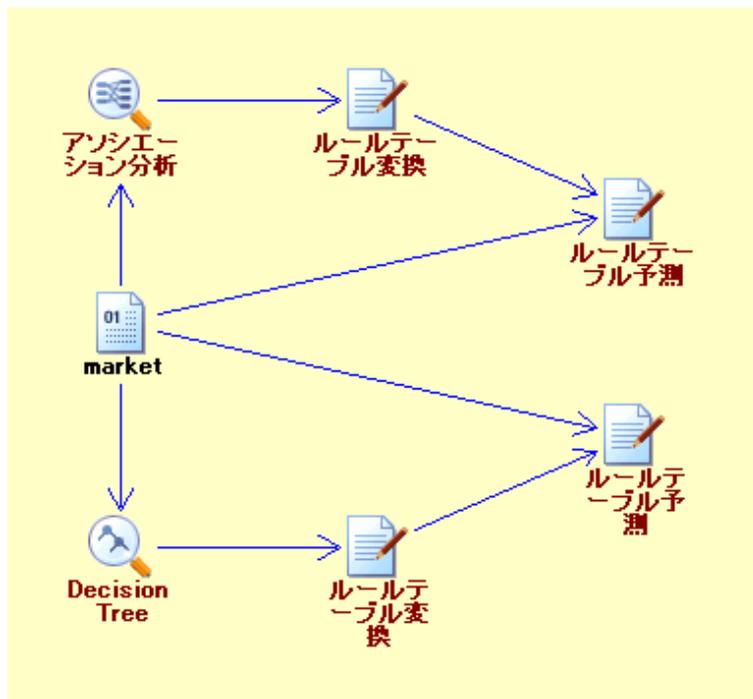
(イ) 分岐変数の外部指定

決定木モデルの作成時に、分岐変数を外部指定することにより、強制分岐させることが可能になり、より柔軟なモデルを作成できるようになりました。



6. Rule Base システム

ルールを通常のテーブル形式で表します。このため、モデルを手動で修正したり、ルールの追加などを簡単に行うことができます。ルールベースシステムを通じ、データマイニングの手法と従来のエキスパートシステムとを融合することにより、新しい分野への進出が期待されます。



アソシエーションルールと決定木ルールの両方をルールテーブルに変換し、予測に使うことができます。

ルールテーブルのサンプル

rulebase - データシートビュー

ファイル(F) 編集(E) オプション(O)

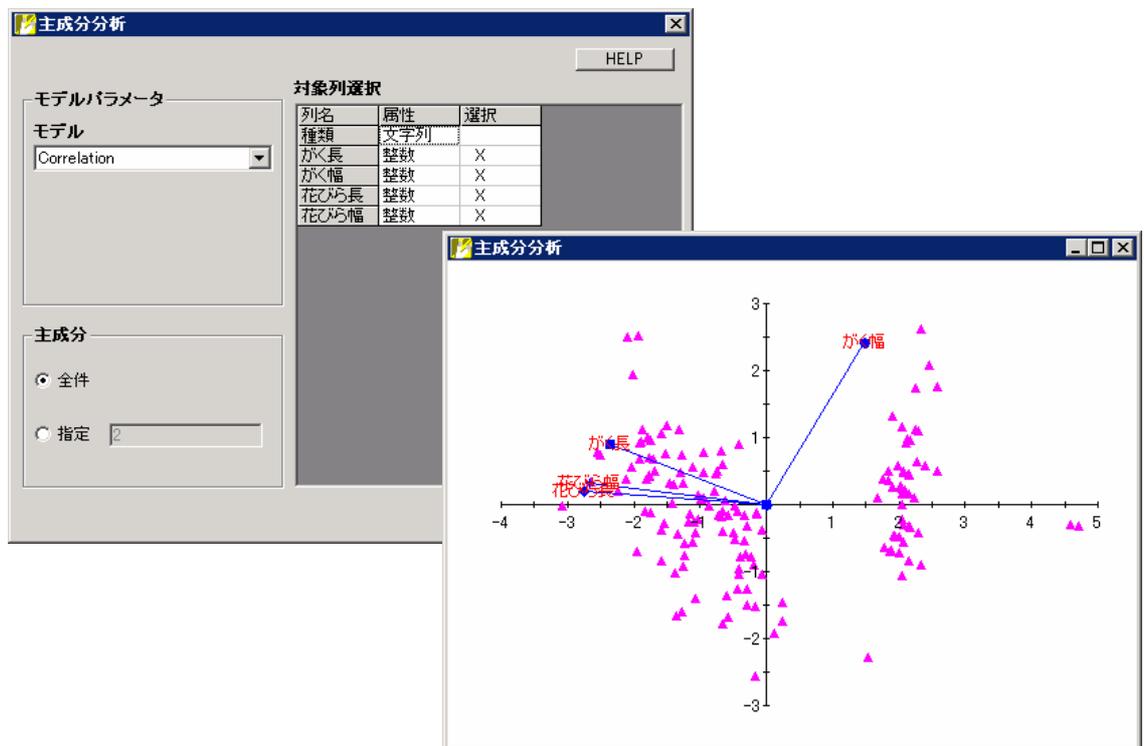
result

Rule No	列名	パターン	数値From	数値To	結論	確信度	サポート
1	1 card	NA	0.000	0.000	pick-OCC	100.000	1.700
2	2 reach.out	NA	0.000	0.000	pick-OCC	100.000	1.700
3	3 moves	5	0.000	0.000	pick-ATT	91.667	1.100
4	4 usage	25	0.000	0.000	pick-ATT	78.571	1.100
5	5 reach.out	Y	0.000	0.000	pick-ATT	77.419	4.800
6	6 employment	S	0.000	0.000	pick-ATT	76.923	1.000
7	7 usage	19	0.000	0.000	pick-ATT	75.000	1.200
8	8 moves	4	0.000	0.000	pick-ATT	75.000	1.200
9	9 usage	15	0.000	0.000	pick-ATT	73.333	1.100
10	10 employment	U	0.000	0.000	pick-ATT	69.991	1.000

7. 統計解析機能

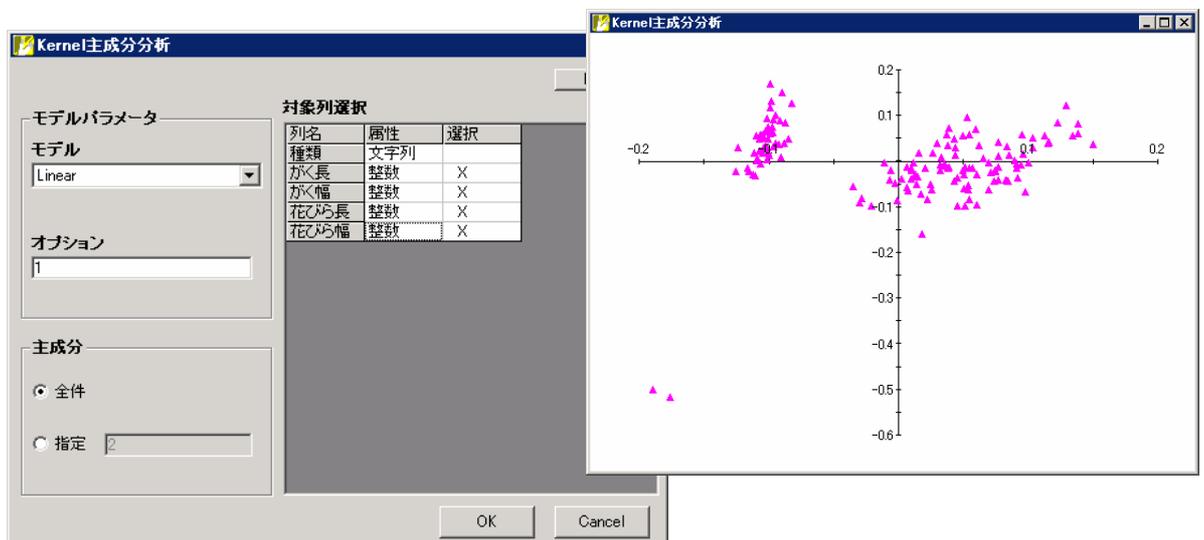
(ア) 主成分分析

代表的な多変量解析の手法の一つである主成分分析が実行できるようになりました。これまでは S-PLUS の機能呼び出すのみでしたが、Visual Mining Studio の機能として実装することにより、大量のデータをより高速かつ安定的に処理することができます。



(イ) カーネル主成分分析

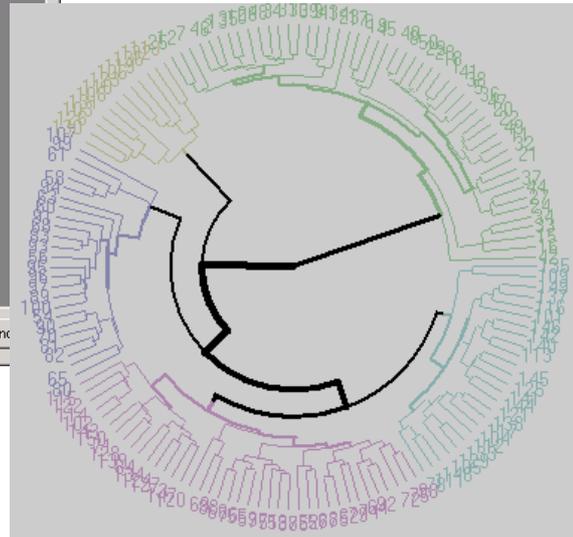
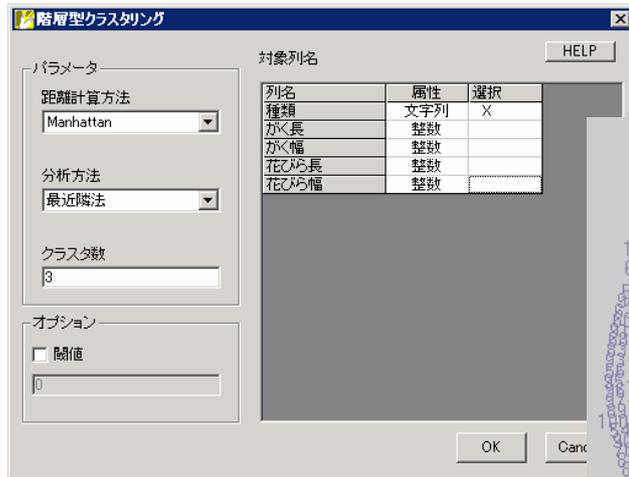
非線形に拡張した主成分分析であるカーネル主成分分析を実行できます。



8. 階層型クラスタリング

擬集型の階層型クラスタリングを行い、デンドログラムを出力します。

クラスタの生成方法として、最近接法、Ward 法などの標準的のものから選択でき、デンドログラムの描画方法は、トーナメント形式のものや、円の外側に広がる描画などが可能です。



9. スクリプト関数の追加

文字列照合関数「belong」

テキストファイルの処理ライブラリ（関数群）

日付/時刻関数

オブジェクト単位のプリント機能

モデルの入出力機能

列丸ごと置き換え関数

計算時間計上関数

その他