

Text Mining Studio 2.0

## 新機能紹介

(株)数理システム

小木しのぶ

# Text Mining Studio とは

---

- さまざまなテキストから知識を得るためのテキストマイニングツール
- 適用例
  - アンケート分析
  - コールセンターオペレーターメモ分析
  - 特許・論文分析
  - Web掲示板分析
  - 営業日報分析

# Text Mining Studio の特徴

---

- 豊富な分析機能
- 豊富なグラフ機能
- 文節単位での分かち書き
- 自由度の高い属性加工
- 分析結果から簡単原文参照
- HTMLレポート作成機能

# Text Mining Studio 2.0

## 新機能

---

- グラフ表示の大幅機能アップ
- 類義語自動抽出
- テキストの自動カテゴリ化機能
- データマイニングとの完全連携
- 分析系でさまざまな改良

他

# グラフ表示の大幅機能アップ

---

- 表示可能なグラフ数が大幅に増加
- 凡例やグラフの位置を自由に調節可能
- 表示軸を自由に選択可能
- 結果表との連動(データソートなど)
- 部分表示(部分拡大)
- 立体グラフのライティングや表示微調整などが可能に
- グラフからの原文参照や辞書登録が可能に

# 類義語自動抽出

- 文字のあわせこみによる類義語抽出
  - アルファベットの大文字と小文字 (MSIとmsi)
  - 漢数字と英数字 (2千5百と2500)
  - カンマのある数値とない数値 (1,000 と1000)
  - 「 - (マイナス)」「 (ハイフン)」「 (ダッシュ)」「ー (長音)」
  - カタカナの最後の長音有無 (サーバーとサーバ)
- 使われ方による類義語抽出
  - 係り受けによる類義利用
  - 共起による類義利用

# テキストの自動カテゴリ化

---

- 表現が異なっても意味的に同じグループをカテゴリという
- カテゴリルールを分析から自動生成
  - ことばネットワーク
  - 文章分類
  - グループピング
    - グループの自動抽出機能
- カテゴリルールからテキストをカテゴリ化し、そこからさまざまな分析を

# データマイニングとの完全連携

---

- 汎用データマイニングツールVisual Mining Studio 上でText Mining Studioが動作
- 従来のTMS for VMStudio (現テキストモジュール) は分析機能が存在しない
- Text Mining Studio上で行った全ての分析結果をシームレスにデータマイニングで利用可能
  - 一般的な分析はText Mining Studio上で
  - カスタムな分析はVisual Mining Studio上で

# 分析系 改良

- 同一行の単語および係り受けの頻度を1にカウント可能に
- ことばネットワークや文章分類での削除語の指定が可能に
- 特徴分析での特徴語一覧の表示可能に
- 述語属性に「すぎる」および複数の組み合わせ「したくない」「できてほしい」などが追加
- 対応バブル分析で最大抽出軸数まで結果を抽出可能になり、グラフでも任意の軸で表示可能に

# 今後

---

- 予測機能
  - テキストを用いた予測の精度向上を実感していただきたい
- 時系列分析を充実
- 実際の分析例を充実
- 使い勝手の向上
- 分析精度の向上