

# データのキモ、見つけます

～特徴的な構造を抽出するデータマイニング技術～

## どんな研究?

近年のビッグデータ解析を語るうえで、統計的なデータマイニング手法の存在は無視できません。この展示では、統計的機械学習技術により、データに潜むパターンや関係性の中心となるコミュニティなど、少数の特別な要素の関係としてデータを構造化・説明する技術を紹介します。

## どこが凄い?

事前に人手によるルール設定なしで、統計的機械学習手法により自動的に複雑なデータを少数の重要な要素に分解して、ユーザに分かり易い構造を抽出します。更に、必要なパターン数をデータの内容に合わせて推定したり、データの理解に不要なノイズ情報に惑わされない手法を開発しました。

## どんな風に役立つ?

人手による解析が困難な大量・大規模なデータ解析に利用することが出来ます。例えば顧客とアイテムの購買履歴から顧客のグループと好みの商品群を推定する、社内のメール履歴やツイートの履歴から社員の派閥や仲良しユーザのコミュニティなどを抽出するなどが可能になります。

### 統計的機械学習による自動的なデータ構造解析

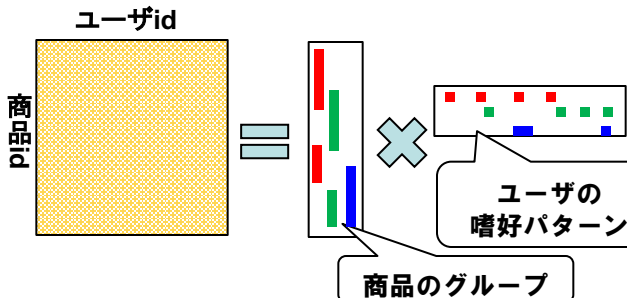
$$p(M|X) \propto p(X|M)p(M)$$

マイニングした構造の評価

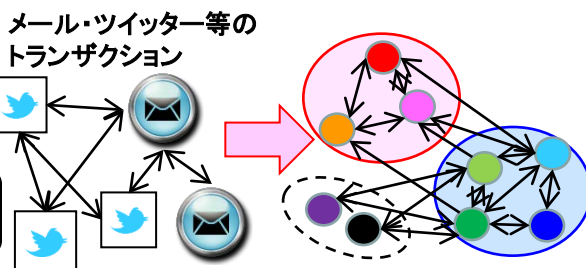
データ適合度

構造の“モデル”

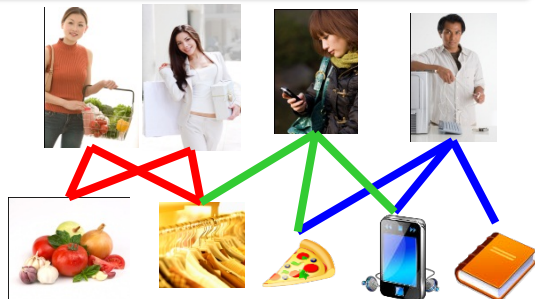
#### Matrix Factorization, トピックモデル: 「このパターンだと、この反応」



#### 無限関係モデル: オブジェクトの「つながり・かたまり」



#### 顧客グループと好みの商品群の同時推定



#### 仲良しグループ等の関係コミュニティ抽出



#### 関連文献

- [1] K. Ishiguro, N. Ueda and H. Sawada, "Subset Infinite Relational Models," Proc. AISTATS, 2012.
- [2] K. Ishiguro, T. Iwata, N. Ueda and J. Tenenbaum, "Dynamic Infinite Relational Model for Time-varying Relational Data Analysis", Proc. NIPS, 2010.
- [3] T. Iwata, T. Yamada, Y. Sakurai and N. Ueda, "Online Multiscale Dynamic Topic Models," Proc. KDD, 2010.

#### 連絡先

石黒勝彦 (Katsuhiko Ishiguro) 協創情報研究部 知能創発環境研究グループ  
E-mail : ishiguro.katsuhiko[at]lab.ntt.co.jp ( {at} の部分を @ に置き換えてください)