

03

ビッグデータ解析から無限データ解析へ ～無限次元行列の確率モデルを用いたデータ解析～

どんな研究

関係データ（SNSでの人のつながりや、ユーザの商品の購買履歴など）から**長方形分割模様**の関係を抽出する研究です。ビッグデータと呼ばれるようなユーザの数や商品の増加とのイタチごっこを本質的に解消するために、潜在的に**無限のデータ**の解析手法の創出を行っています。

どこが凄い

任意の（無限通りの）**長方形分割模様**を表現可能なモデル(Gilbert tessellation)は古くから知られていましたが、その統計的な性質の解析は極めて難しく、実際のデータ解析に用いることは出来ていませんでした。本研究では**解析に適した新しい表現モデル**を考案しました。

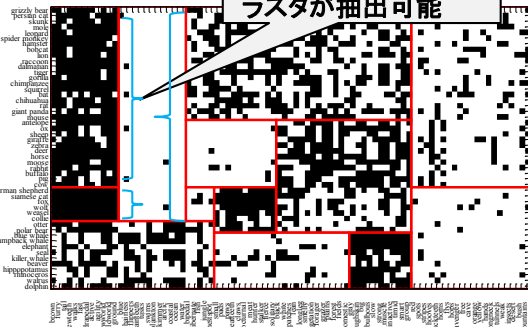
目指す未来

関係データに対するビッグデータ解析においては、増え続けるデータに対応するだけでなく、多種多様なあらゆる関係性に対応していくことが重要になります。現在、多くのデータ解析システムはまだ人間の介入を必要としています。我々の取り組みは**人間の手を離れた自立的な解析システムの確立**に繋がります。

関係データ解析

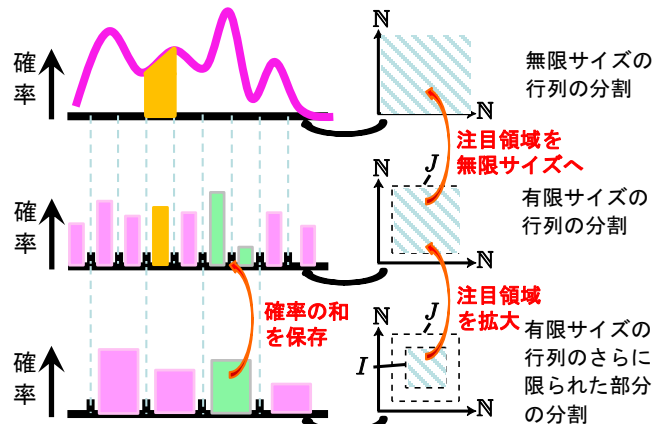
動物(55種類) × 属性(80種類)のバイナリデータから長方形分割クラスタを抽出した解析例：

属性や動物において多重の意味を持つクラスタが抽出可能



確率モデルの無限拡張

有限モデルの**極限**として無限モデルを構成。



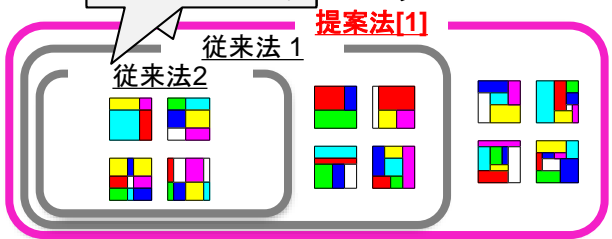
確率モデルを用いた関係データ解析

確率モデルに基づく解析の基本方針：確率モデルを観測データにフィッティング（**対象を表現する確率モデルの設計が重要**）。

本研究の貢献：無限サイズの行列における**全ての（無限通りの）長方形分割**を表す確率モデルの構成に成功。従来は限定された長方形分割が主として扱われてきた。

従来は極めて限られた形の長方形分割しか抽出することが出来なかった。

あらゆる長方形分割を扱うことが出来るため、**従来見つけることの出来なかった関係性**が新たに見出せる可能性がある。



関連文献

[1] M. Nakano, K. Ishiguro, A. Kimura, T. Yamada, N. Ueda, "Rectangular tiling process," in *Proc. International Conference on Machine Learning (ICML)*, 2014.

連絡先

中野允裕 (Masahiro Nakano) メディア情報研究部 メディア認識研究グループ
E-mail: nakano.masahiro(at)lab.ntt.co.jp

