

大事な特徴を捉えて流れるデータを把握する

～多次元時系列データからの特徴抽出～

どんな研究

本研究は、多次元（多くの特徴量を持つ）時系列データから時間的なパターンを抽出する技術です。時系列データの連続性の制約のもとで、各次元の特徴の有効性を推定して、時系列データに内在する特徴の共起パターンを確率モデルで推定します。

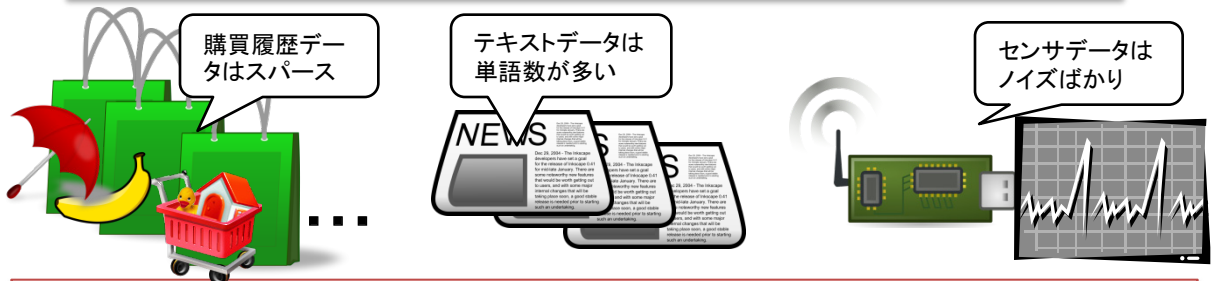
どこが凄い

本技術は、時系列データからパターンを見つけるだけでなく、「パターンを特徴づける特徴量」も自動的に抜き出します。これにより、大量のノイズが混ざった時系列データからもパターンを発見できる上に、解析後のパターンの解釈などの人手の工数を減らすことに貢献できます。

めざす未来

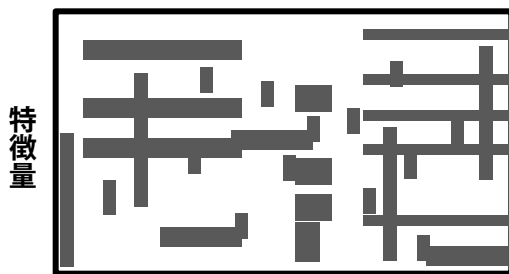
「ビッグデータ」に代表される多次元のデータから隠れた関係や知識を発見する技術は、今後ビジネスプロセスの効率化や業績向上のため益々必要となります。我々の研究技術は、実用的な精度と時間で「ビッグデータ」を活用できる社会の実現に役立ちます。

多次元の時系列データの全ての特徴量が本当に重要なのか？



この技術は時間的なパターンの違いを“特徴づける”特徴量だけを抽出

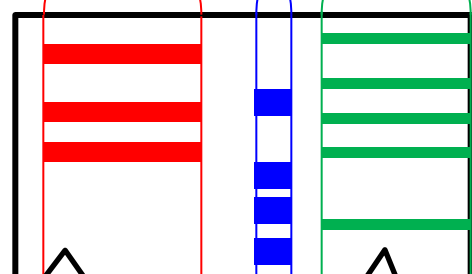
観測される時系列データ



時間

⊖ 何が何やら・・・

パターン1 パターン2 パターン3



時間

☺ ここにパターンがあったのか！

☺ このパターンはこの特徴量が大事なのか！

- ☺ 大量のノイズに影響されずに時間パターンを見つける
- ☺ 各パターンに重要な特徴量が紐づけられ、解析結果の解釈が容易になる
- ☺ パターンの数を自動的に推定

関連文献

[1] 石黒勝彦, 上田修功, 澤田宏, “動的サブセットクラスタリング,” 第14回情報論的ワークショップ(IBIS), 2011.

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}の部分をも@に置き換えてください)