

02

データに潜む本質的パターンを素速く見出す

～多次元複合データ分析の高度化・高速化～

どんな研究

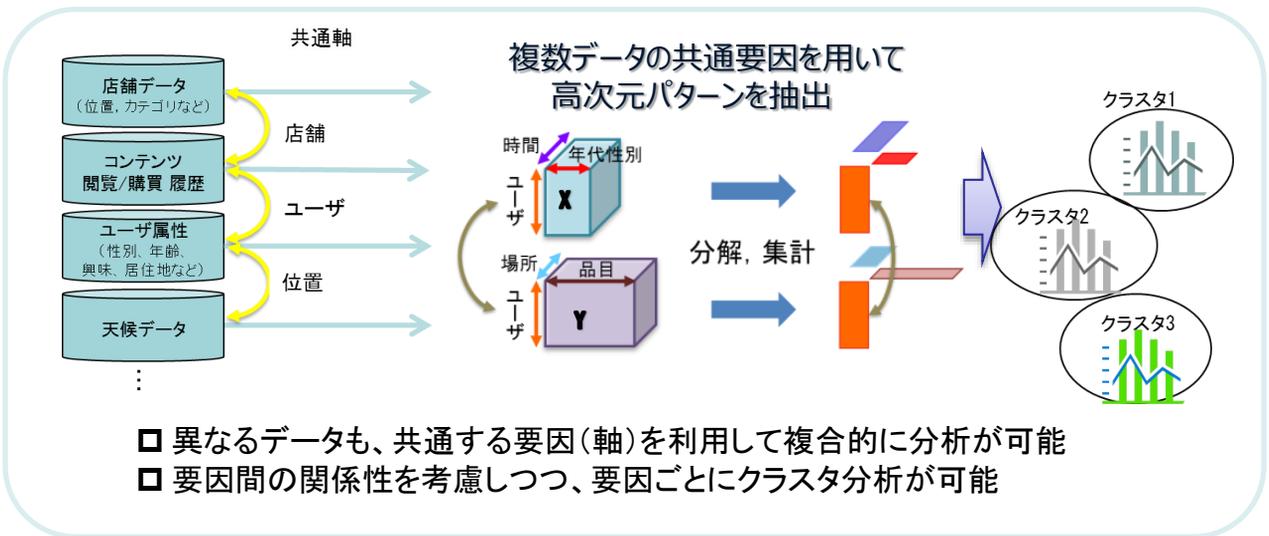
複数の要因を同時に分析できる技術の研究をしています。例えば、ユーザ・ユーザ属性・購入品・購入時間・購入店舗など、複数要因を同時に考慮することにより「いつごろ、どんな店で、どんな人たちが、何を購入する傾向があるか」という分析をより詳細にかつ高速にできます。

どこが凄い

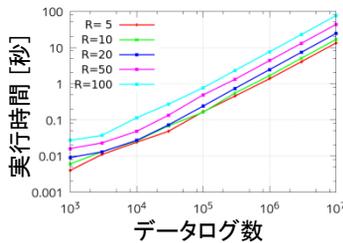
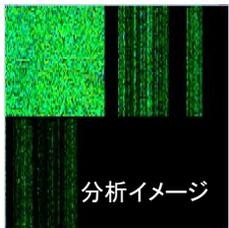
高速性：データの疎密を考慮したメモリ通信とアルゴリズムの最適化によって処理の高速化を達成。数億レコードも数時間以内に処理可能です。
整合性：要因間の属性情報を紐付けする制約項を導入し、一部の情報が欠損したデータでも、整合性を保持した分析が可能です。

目指す未来

現在、機械や人間行動のログとして莫大な情報量が蓄積され、分析技術のさらなる高速・高度化に対するニーズが高まっています。本技術は、今後増え続ける多くの**超大規模なデータ**を**データ間の関連を深く結びつけ**ながら、これまで知り得なかった因果関係や知識の発見を促します。

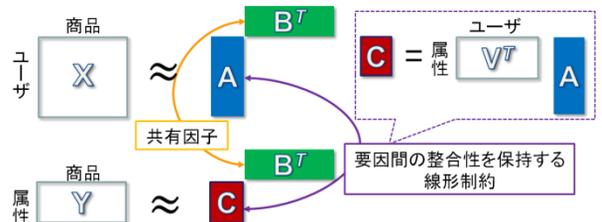


高速化



- ・スパース性を考慮した超高速アルゴリズム
- ・1000万ログ(基底数10、反復100回)の非負値スパーステンソル因子分解も30分で可能

高度化



線形制約項によってデータ間の分布を考慮した因子分解手法の様々な拡張

関連文献

- [1] 松林達史, 幸島匡宏, 林亜紀, 澤田宏, "非負値テンソル因子分解を用いた購買行動におけるブランド選択分析," 人工知能学会論文誌, 2015.
- [2] 幸島匡宏, 松林達史, 澤田宏, "異粒度データ解析のための非負値行列分解手法," 2015年度人工知能学会全国大会, 2015.
- [3] 幸島匡宏, 松林達史, 澤田宏, "異粒度複数行列の制約付き非負値因子分解について," 第17回情報論的学習理論ワークショップ, 2014.

連絡先

松林達史 (Tatsushi Matsubayashi) サービスエボリューション研究所
 E-mail: matsubayashi.tatsushi(at)lab.ntt.co.jp

