

## ソーシャルビッグデータのリアルタイム蓄積・解析基盤の開発

ソーシャルビッグデータ研究連携センターでは、ソーシャルビッグデータのリアルタイム蓄積・解析基盤の開発を目指し、1. ソーシャルメディアにおけるユーザ間のつながりを表すソーシャルグラフや、2. 時刻情報を含むトランザクションデータ等の非テキストデータに関する高度データマイニング技術に関して研究開発を行った。また、3. ソーシャルメディアが人々の行動に与える影響分析技術及び4. ソーシャルメディアにおける時空間情報に着目した大規模情報統合可視化技術の研究開発を推進している。さらに、ビッグデータ活用研究室と連携し、ソーシャルビッグデータ連携による環境リスク分析と行動支援技術の開発・実証を推進している。

## ■平成28年度の成果

## 1. ソーシャルグラフに対する高度マイニング技術開発

グラフデータマイニングに関しては、クラウド環境に適したスケーラブルな分散グラフデータベースエンジンの開発を行った。多くのグラフデータは次数分布に偏りがあり、通常の分散グラフデータベースエンジンでは効率的な処理が難しい。

平成28年度は、前年度までに開発したGraphSlice手法に対して、分散計算における通信コストを最小化することによる処理最適化の検討を行った。図1に示すように最適化問題を2部グラフの最大マッチング問題と同等に考えることで、最適な処理を提供する。提案手法をApache Sparkで実装し、通信コストが平均で12%削減されることを確認した。

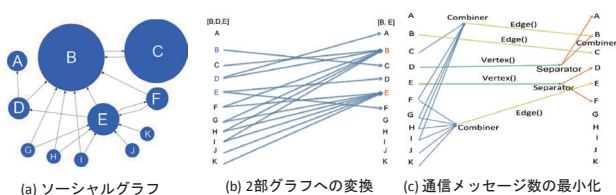


図1 GraphSliceにおける通信コストを最小化することによる処理最適化

## 2. 希少なアイテムに関する周期的頻出パターンマイニング技術開発

ビッグデータから希少なアイテムに関する知識を獲得することは重要な課題の1つである。従来の頻出パターンマイニング手法では、アイテムの希少性が考慮されていないため、無用なパターンが多数抽出されてしまうという問題が存在する。

平成28年度は、この問題を解決するための周期的頻出パターンモデルを提案した。提案モデルは、periodic-all-confidenceという新たな指標に基づき、アイテム間の周期の相関性が高いパターンを選択することで、無用なパターンが出力されることを防ぐ。さらに、周期的頻出パターンの効率的なマイニングアルゴリズムであるEnhanced Periodic-Frequent Pattern-growth (EPF)を開発した。EPFはトランザクションデータから変換された木構造データを再帰的に探索することで提案モデルの要件を満たすすべてのパターンを列挙する(図2)。人工データ及び現実データの双方において、提案手法が既存の最新手法と同等あるいはそれ以下の実行時間で希少なアイテムに関する周期的頻出パターンを発見可能であることを確認した(図3)。

## 3. ソーシャルメディア影響分析技術開発

ソーシャルメディア上での他者との対話や投稿の閲覧は、オンラインだけでなく実世界にも影響を与える。ソーシャルメディアの影響範囲の解明及び社会生活における意思決定や行動選択の支援を目的として、実世界での人々の行動を変化させるソーシャルメディア情報の検索及び分析技術の研究開発している。

平成28年度は、実世界での主要な人間行動の1つで

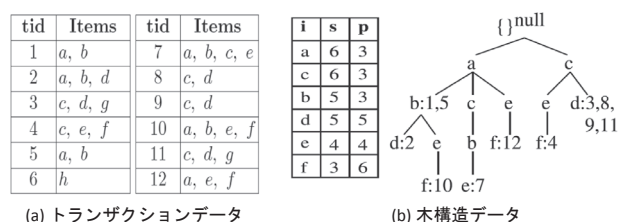


図2 データ構造の変換に基づく効率的な周期的頻出パターンの発見

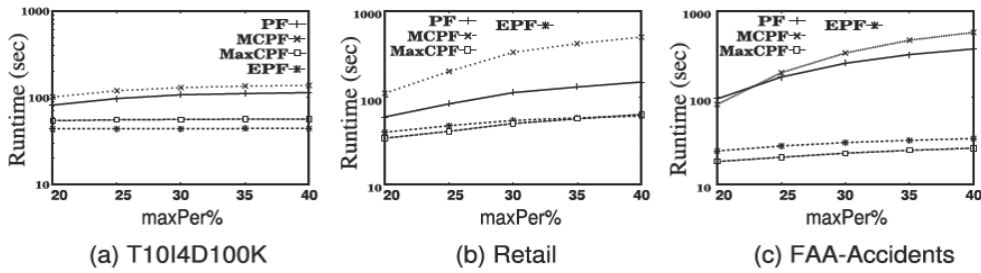


図3 提案手法EPFと既存の最新手法との実行速度の比較

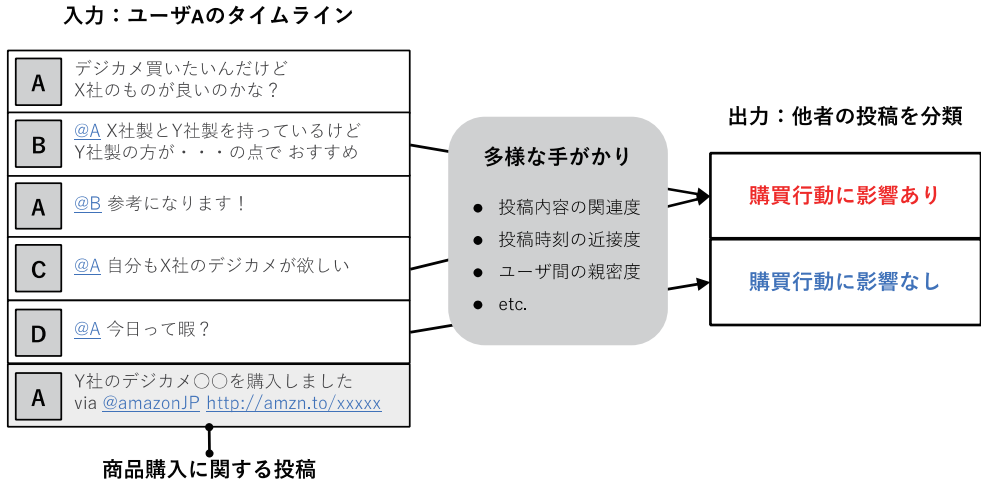


図4 購買行動に影響を与えたソーシャルメディア情報の発見

ある「購買」を対象として、ソーシャルメディア上から人々の購買行動に影響を与えた投稿を発見する手法を提案した。提案手法は、投稿内容の関連度・投稿時刻の近接度・ユーザ間の親密度などの様々な手がかりを用いて、購買者のタイムラインから商品購入に影響を与えた他者の投稿を検出する（図4）。1年間のTwitterデータに対して、提案手法はF値0.53という精度で影響力の高い投稿を検出することに成功した。

#### 4. 大規模情報統合可視化技術の研究開発

ソーシャルビッグデータとして得られるテキスト・非テキストデータより、実空間で日々発生するイベントの影響や時空間的な広がり理解可能とする3次元可視化手法の研究開発を行った。マイクロブログストリーム中の位置参照表現に着目し、位置座標情報が付加されていないつぶやき(Tweet)についても、地名や施設名などの位置参照表現に基づいて位置に関連付け、局所的なイベント、広範囲イベントを認識して3次元空間に多層的に可視化し、その時間変化をアニメーション可能と

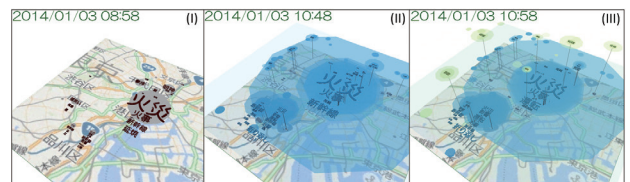


図5 時系列トピック3次元可視化統合基盤フレームワーク

する可視化システムを実現した（図5）。

平成28年度は、具体的に火災やゲリラ豪雨などの災害及び事故発生時の状況可視化を目的として、雨量情報や物理センサーと連携可能にする統合フレームワークを提案した。物理世界とバーチャル世界を相補的に探索することで、個別のセンサーデータからでは分からない、イベントの要因、関係性などが見えてくる。図5はJRの運行に大きな影響を与えた火災発生時の状況を可視化した事例である。ゲリラ豪雨時の状況可視化については、ビッグデータ活用研究室において行われているリスクを考慮したドライブ経路探索への適用を目指して連携を行っている。