3.10.6.2 ソーシャルビッグデータ研究連携センター

連携センター長（兼務）　木俵　豊　ほか5名

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

■概要

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．時刻情報付きトランザクション集合からなる時制データベースに対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ている。また、３．ソーシャルメディアが人々の行動に与える影響の分析技術及び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、ビッグデータ利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。

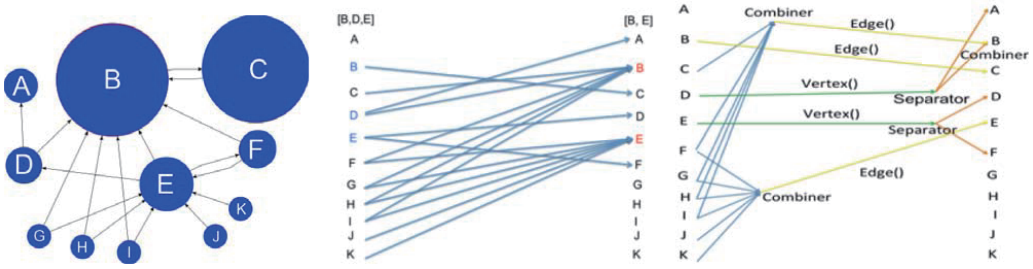
■平成29年度の成果

１．ソーシャルメディアに対する高度グラフマイニング技術開発

我々はこれまでに、ソーシャルグラフに対する効率的な分散処理フレームワークであるGraphSliceを提案してきた。平成29年度は、GraphSliceにおけるグラフ処理計画の最適化に関する研究を実施した。提案手法は、ソーシャルグラフ（図１（a））を、グラフ処理における通信パターンに関して、それと等価な２部グラフに変換する（図１（b））。次に、２部グラフにおける最小頂点被覆問題を解くことで、最適なグラフ処理計画を発見する（図１（c））。提案手法をApache Spark上に実装し、通信コストが平均で12％減少することを確認した。また、実用的なグラフマイニングタスクとして、テキスト中のあいまいな言及に対応するエンティティを所与のリストから発見するList-only Entity Linking（List-only EL）に取り組んだ。List-only ELは、知識ベースには含まれにくい新製品や希少イベントに関する情報をソーシャルメディアから発見するうえで重要な役割を担う。

２．時制データベースに対する部分周期的パターンマイニング技術開発

時刻情報付きトランザクション集合により構成される



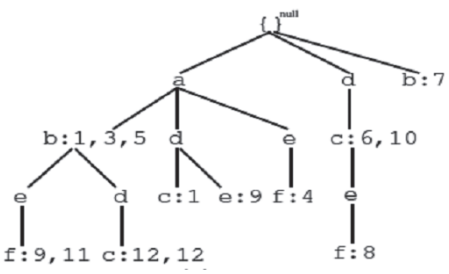
（a）ソーシャルグラフ　（b）等価な2部グラフ　（c）ノード間の通信数の最･化

図1　2部グラフへの変換と最小頂点被覆問題への還元に基づくソーシャルグラフ分散処理の最適化

tid	*ts*	*items*
101	1	*abg*
102	1	*acd*
103	3	*ab*
104	4	*aef*
105	5	*abg*
106	6	*cd*
107	7	*bg*

tid	*ts*	*items*
108	8	*cdef*
109	9	*abef*
110	9	*ade*
111	10	*cdg*
112	11	*abef*
113	12	*abcd*
114	12	*abcd*

i	pf
a	8
b	7
d	5
c	4
e	3
f	2



（a）時制データベース　（b）木構造データ

図2　データ構造の変換に基づく時制データベースからの効率的な部分周期的アイテム集合の発見

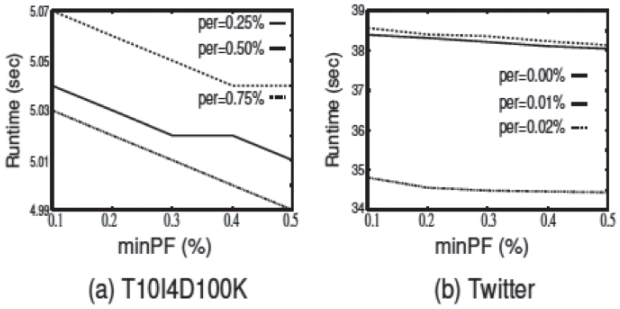


図3　部分周期的アイテム集合の発見アルゴリズム3P-growthの実行時間の検証

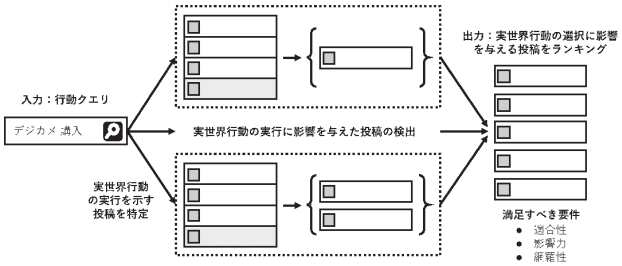


図4　実世界行動の選択に影響を与える投稿をソーシャルメディアから検索

時制データベースから、部分的な周期性を持つアイテム集合を発見することは、実世界における購買や事故等のパターンに関する知識を獲得するうえで重要な研究課題である。平成29年度は、全期間における周期的な出現頻度に基づきアイテム集合の周期性を測る指標としてperiodic-frequencyを提案した。さらに、時制データベースから変換された木構造データを再帰的に探索し、部分周期的な全てのアイテム集合を効率的に発見するPartial Periodic Pattern-growth（3P-growth）なるアルゴリズムを開発した（図２）。人工データ及び現実データ（Twitter）のそれぞれにおいて提案手法の計算時間が十分に短いことを示した（図３）。また、Twitterデータに対する実験結果の観察を通じて、実世界のイベントに関するキーワードを発見可能であることを確認した。

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並びに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を研究開発している。平成29年度は、実世界での主要な人間行動のひとつである「購買」を対象として、ソーシャルメディアから人々の購買行動の選択に影響を与える投稿を検索する手法を提案した。提案手法は、行動の種類を指定する所与のクエリに対して、クエリに関連する行動の実行に影響を与えたソーシャルメディア上の投稿集合を検出する。次に、得られた投稿集合を適合性・影響力・網羅性の観点からランキングすることで行動選択の判断に有用なものを上位に配置する（図４）。1 年間のTwitterデータを用いた評価実験によって、提案手法が影響力のある多様な投稿集合を検索可能であることを確認した。

４．大規模情報統合可視化技術の研究開発

Twitterなどのマイクロブログ記事の位置参照表現を利用し、投稿中の各単語の時空間的な局所性を単位領域ごとに算出し、これらをワードクラウド表現により地理空間中に可視化する手法を開発してきたが、本年度はこの技術を応用し、様々なセンサデータとの統合可視化を行うための基本技術開発を進めた。具体的には、ビッグデータ利活用研究室と連携し、ソーシャルビッグデータからゲリラ豪雨の発生に伴う交通や人々の反応の変化を抽出及び可視化することで豪雨リスクをより的確に把握するための技術を開発した。ソーシャルメディアから得られた豪雨による影響情報の地理空間ワードクラウドと、PANDAレーダから得た豪雨の警戒円及びXRAINから得た実際の降雨状況の可視化を統合可視化し、実際に台風やゲリラ豪雨が発生した場所日時を事例として用いたプロトタイプを実現した（図５）。

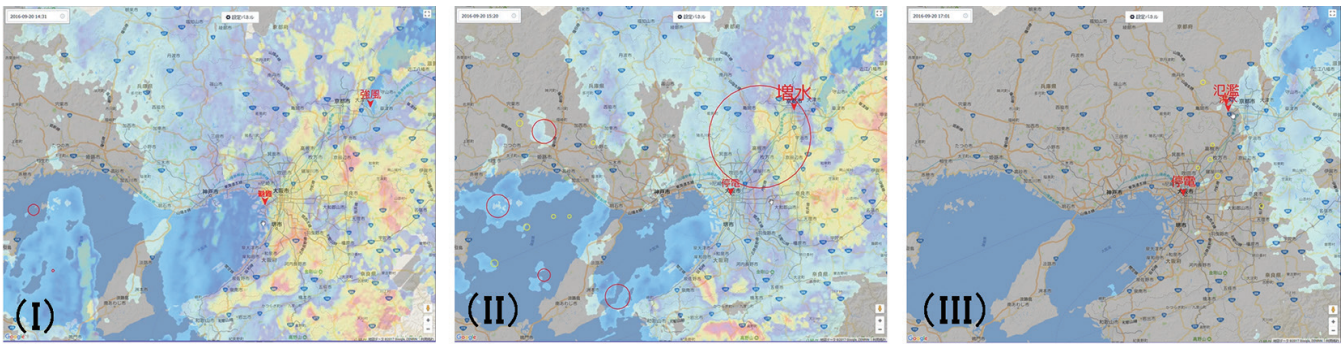


図5　豪雨データ（PANDA、XRAIN）とソーシャルビッグデータ地理空間ワードクラウドとの統合可視化