

## 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・解析基盤の研究開発を行っている。平成30年度は、ソーシャルメディアデータを用いてユーザ間のつながり（ソーシャルグラフ）や人々の行動に与える影響を分析する技術と、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た。さらに、ビッグデータ利活用研究室が開発する異分野データ連携プラットフォームと連携し、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や、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実装に取り組んだ。

## ■平成30年度の成果

## 1. 高頻度かつ高効用なアイテム集合の効率的な発見方式の開発

データベースの中から効用の高いアイテム集合を発見すること（表1）は、売上の高い商品の組合せの発見や、長い渋滞が発生しがちな経路の発見など、実世界における様々な応用が存在する重要な問題である。高効用アイテム集合の発見に関する従来研究では、出現頻度が考慮されていなかったため、興味のない低頻度なアイテム集合まで発見されてしまう。また、従来手法は計算コストが高く実用に不向きであった。これらの問題を解決するため、平成30年度は、事故や災害などまれなケースにおいて相関性の高いデータを発見する処理を高速化するアルゴリズム（Sequential/ Parallel Weighted FP-growth）を開発し、無用な相関パターンの枝刈りやパターン発見の終了条件を工夫することで、実行時間とメモリ消費量を従来（WFI方式）の約半分にまで削減することに成功した。この成果をビッグデータ分野のトップカンファレンスであるIEEE Big Data 2018国際会議（採択率19%）等で発表した。また、出現頻度及び効用に関するユーザ指定の閾値を上回るアイテム集合を効率的に発見可能な手法であるHigh Utility Frequent Itemset Miner（HU-FIMi）を開発した。提案手法は、cutoff utility及びsuffix

utilityという2つの枝刈り指標を考慮することで、候補となるアイテム集合の探索空間を削減する。購買履歴に関する実データ及び人口データを用いた評価実験の結果、提案手法（HU-FIMi）は既存手法（EFIM）に比べて、メモリ使用量と同程度に保ったままデータベースのスキャン回数を削減することで実行時間を大幅に短縮可能であることが確認された（図1）。さらに、種々の閾値に対する提案手法の出力の変化の分析を通じて、アイテム集合の出現頻度と効用の関係性を検証した。本成果は、知識発見及びデータマイニングに関する国際会議PAKDD2019への採択が決まった（採択率24.1%）。

## 2. ソーシャルメディアデータに基づく人間行動解析技術の開発

ソーシャルビッグデータの利活用において、ユーザの興味に基づく情報推薦とは異なる新たな枠組みを実現すべく、平成30年度はユーザのスキルを成長させる情報の提供に関し、「ユーザスキルの成長過程のモデル化」と「アイテムの難易度の推定」という2つの重要な課題に取り組んだ（図2）。第1の課題については、アイテムの選択行動を繰り返すことによるスキルの成長を潜在変数として学習する進行モデルを提案した。第2の課題については、「アイテムの選択は現在のスキルに依存して行われる」という仮定に基づき、進行モデルの学習結果を利用して各アイテムの難易度を推定する統計的な手法を提案した。スキルアップ推薦システムへの応用に向けて、レビューサイトや添削サイト等のソーシャルメディアから取得した実データ4種類及びユーザスキルとアイテム難易度の正解を含む人口データ1種類を用いて評価実験を行った。その結果、スキルの大小に応じて頻繁に選択されるアイテムの傾向が異なること（表2）、並びに複数のアイテム属性を考慮することでユーザスキル及びアイテム難易度をより正確に推定できることを確認した。

また、ソーシャルメディアデータを用いた人流解析の研究開発にも着手した。実世界における人々の動きを把握することは、イベント分析や経路推薦など多様な応用の可能性を秘めている。本研究では、ソーシャルメディ

アに投稿されたリアルタイム情報と、移動体に関する統計情報を組み合わせて解析することで、人流や混雑をより正確に予測することや、変化の要因を説明することを目標としている。平成30年度は、次年度以降におけるこれらの課題の達成に向けて、必要なデータセットの整備や候補アプローチの調査・検討を行った。今後、ビッ

グデータ利活用研究室で開発されている異分野データ連携プラットフォームや委託研究（課題201）における異分野データ連携によるスマートモビリティ基盤の研究開発と連携し、スマートで持続可能なモビリティサービス（Smart Sustainable Mobility）の応用実証に資する基盤技術の開発を進める。

表1 時刻情報付きトランザクション集合の例（購買履歴に関するデータベース（左）と各アイテムの効用（右））

tid	Items	tid	Items
1	(a, 2), (b, 3), (f, 2)	5	(a, 1), (b, 2), (c, 1)
2	(a, 2), (c, 1), (d, 3), (e, 2)		(d, 4), (g, 2)
3	(a, 3), (b, 1), (h, 2)	6	(c, 3), (d, 2), (f, 3), (e, 1)
4	(c, 2), (d, 3), (e, 1)	7	(b, 3), (d, 4)

Item	price	Item	price
a	200	e	200
b	300	f	500
c	200	g	200
d	400	h	300

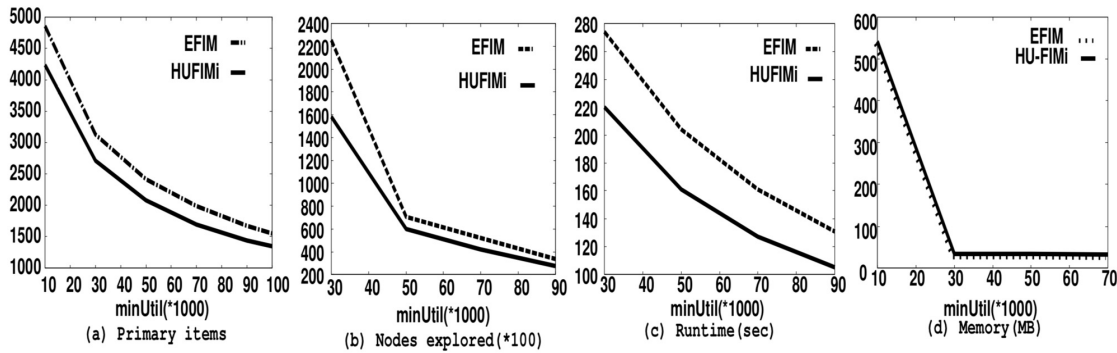


図1 提案手法（HU-FIMI）及び既存手法（EFIM）の計算コストに関する比較

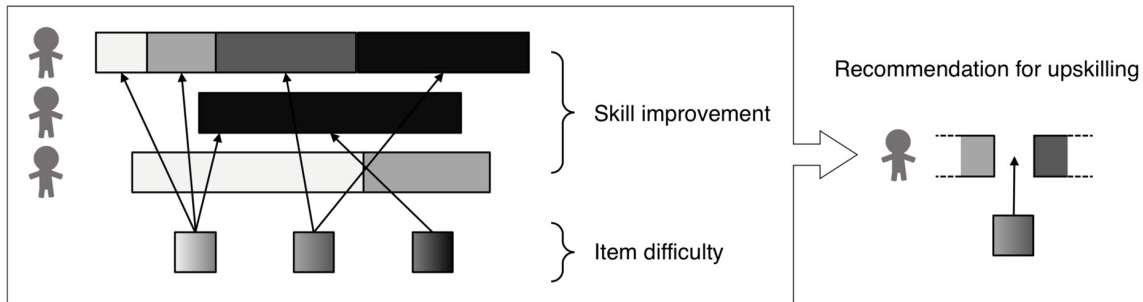


図2 スキルアップのための情報推薦とその技術的課題

表2 スキルの違いに基づく情報推薦の例（映画の選択に与える影響）

(a) Lowest skill level		(b) Highest skill level	
Name	Year	Name	Year
Pulp Fiction	1994	Rear Window	1954
Star Wars: Episode IV - A New Hope	1977	The Sound of Music	1965
Star Wars: Episode VI - Return of the Jedi	1983	The Graduate	1967
Star Wars: Episode V - The Empire Strikes Back	1980	It's a Wonderful Life	1946
Batman	1989	The Birds	1963
Dances with Wolves	1990	Gone with the Wind	1939
Indiana Jones and the Raiders of the Lost Ark	1981	Psycho	1960
The Shawshank Redemption	1994	Casablanca	1942
True Lies	1994	Vertigo	1958
Jurassic Park	1993	Citizen Kane	1941