

**科学研究費助成事業 研究成果報告書**

平成 29 年 5 月 29 日現在

機関番号：25403

研究種目：基盤研究(C)（一般）

研究期間：2014～2016

課題番号：26330139

研究課題名（和文）大規模ソーシャル画像データを用いた動向情報の時空間分析に関する研究

研究課題名（英文）Spatio-Temporal Analysis for Trend Information using Large-Scale Social Images

研究代表者

田村 慶一（Tamura, Keiichi）

広島市立大学・情報科学研究科・准教授

研究者番号：80347616

交付決定額（研究期間全体）：（直接経費） 3,500,000円

研究成果の概要（和文）：近年、ソーシャルメディア上のユーザは日々の出来事をメッセージとともに画像データを通して投稿することで盛んに情報発信をするようになってきている。位置情報とテキストとともに投稿される画像データをソーシャル画像データと呼び、このソーシャル画像データは重要な情報源のひとつとなっている。そこで、ソーシャル画像データを用いた動向情報の時空間分析を可能とするための新しい時空間データマイニング技術が求められている。本研究では、ソーシャル画像データを用いて「何が、いつどこで発生し、どのように変化しているか」を分析可能にするための時空間データマイニングに関する基盤技術の確立を行った。

研究成果の概要（英文）：In these days, people on social media dispatch information by posting messages related to daily activities with image data. Image data including text data and location information are called social image data. They have become one of the most important information source for our daily life. Therefore, new spatio-temporal mining techniques for social image data are required. In this study, basic techniques of spatio-temporal data mining for enabling us to analysis “what and when happened, where is happening, and how it changes.” regarding to daily events and topics using social image data.

研究分野：データ工学

キーワード：ソーシャル画像 時空間マイニング 動向情報 ソーシャルメディア 並列処理

### 1. 研究開始当初の背景

ソーシャルメディアの急速な発達と GPS 付き携帯電話やスマートフォンの普及とともに、インターネット上のユーザは携帯電話やスマートフォンに搭載されたカメラ機能で撮影した写真を画像データとしてソーシャルメディアサイト (Twitter, Flickr や Facebook など) に投稿することでリアルタイムに視覚的な情報発信を盛んに行っている。投稿画像データ数は指数関数的に増加し、ソーシャルメディアサイト上で1日に投稿される画像データ数は5億以上に達している (2013年5月現在)。投稿画像データは個人的な趣味だけでなく社会的な話題や出来事を含み、ビッグデータへの関心の高まりとともにソーシャルメディアサイト上の投稿画像データから有益な知識を発見することが注目されている。

ソーシャルメディアサイト上の投稿画像データは、通常、画像データだけではなく、時間、位置情報やテキストデータ (タイトル、内容記載した文書やタグキーワード) とともに投稿される (図1)。これらの情報は、投稿された画像データの内容記述するための一種のメタデータであり、このメタデータをソーシャルメタデータと呼ぶ。また、ソーシャルメタデータが付与された画像データをソーシャル画像データと呼ぶこととする。ソーシャル画像データはユーザが目にした社会的な話題や出来事を視覚的に伝えることができるため、テキストデータのみでの情報発信を凌駕しつつある。そこで、ソーシャル画像データを対象とした研究が行われているが、テキストデータを対象とした研究と比べると、動向情報の時空間分析に関する基盤技術がまだ確立しているとはいえない。動向情報の時空間分析とは、事象などの変動に関する分析であり、災害時情報分析、マーケティングや流行の把握・予測なども含め、私たちの日常生活において発生する意思決定に重要な役割を果たす。

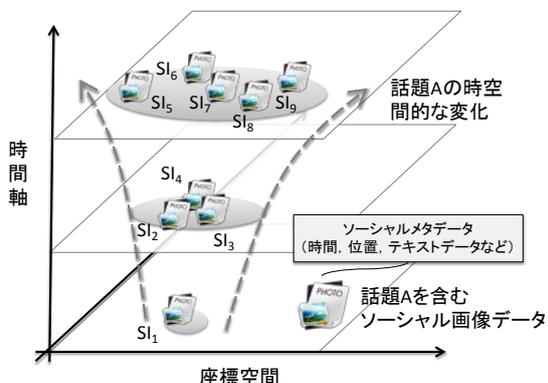


図1 ソーシャル画像データの時空間分析

### 2. 研究の目的

本研究では、ソーシャル画像データを用いた動向情報の時空間分析 (図1) を可能とするための新しい時空間データマイニング技術の確立を目指す。ソーシャル画像データでは画像データの内容だけではなく、画像データとともに投稿されるソーシャルメタデータを考慮した時空間データマイニングの技術が不可欠となる。研究代表者はこれまでに時間と位置情報が付与された文書データに着目し、文書データの時空間データマイニング技術の開発 (若手研究 (B), 課題番号: 23700124) を行なってきた。本研究では文書データの時空間データマイニングを基盤とし、ソーシャル画像データを用いた動向情報の時空間分析への展開を行う。

### 3. 研究の方法

動向情報の時空間分析を行う場合、「何が、いつどこで発生し、どのように変化しているか」を分析可能にするための基盤技術が不可欠である。具体的には、以下の5つの研究課題を取り組み、ソーシャル画像データを対象とした動向情報の時空間分析のための時空間データマイニング技術の確立を行う。

- (1) ソーシャル画像データの時空間データモデル化
- (2) 内容ベースの時空間クラスタリング手法
- (3) 時空間バースト検出手法
- (4) 大規模ソーシャル画像データの効率的な管理法とその並列処理
- (5) 時空間上における動向情報の可視化

また、ソーシャルメディアサイト上で収集した大規模データを用いて評価を行い、実際の動向分析システムのプロトタイプを作成する。

### 4. 研究成果

各項目に対する研究成果は次の通りである。

- (1) ソーシャル画像データの時空間データモデル化

時空間文書データモデルを拡張し、ソーシャル画像データの新しい時空間データモデルを作成した。ソーシャル画像データを時間と位置による時空間情報に加えて、画像データ、テキストデータの特徴量を含めた多次元時空間データとして表現することで、クラスタリングや分類などの機械学習のアルゴリズムを適用することができるようになった。また、ディープラーニングの転移学習モデルを使用した特徴量抽出手法の研究開発も実

施し、典型的な特徴量の抽出手法と比較して、精度の高い特徴を抽出できることを確認した。

## (2) 内容ベースの時空間クラスタリング手法

ソーシャル画像データをその内容ごと分類するための新しい内容ベースの時空間クラスタリング手法を開発した。まず、(1)で開発を行ったデータモデルを使用して、時間と位置に加え、画像データやテキストデータの特徴量を考慮してソーシャル画像の類似度を算出する(図2)。類似度を算出後、密度に基づくクラスタリング手法を応用した時空間クラスタリング手法を用いて、各地域で取り上げられている話題を時空間クラスタとして取り出すことができる。新しい時空間クラスタリングでは、単に密度を基準とするだけでなく、都会と田舎、昼間と夜間など時空間的な密度の差を考慮して時空間クラスタを取り出すことができる。

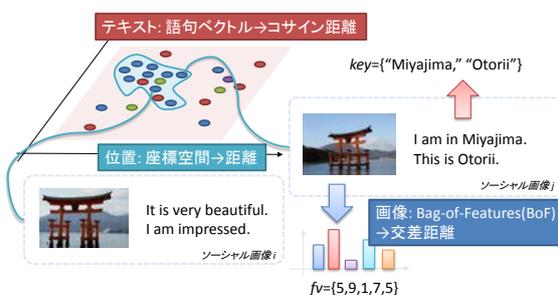


図2 内容ベースの時空間クラスタリング



図3 時空間クラスタリング抽出結果

図3に京都周辺で収集されたTwitter上の画像付きのツイートデータ11,189件のクラスタリング結果を示す。48個の時空間的な話題が時空間クラスタとして抽出でき、テキストのみを使用する場合と比較してより多くの情報を抽出できることを確認できた。これは、テキストのみ同士では類似ではないと判定された投稿について、画像も含めると類似であると判定されたためであり、ソーシャル画像データの時空間データモデルの有効性を確認することができた。

を確認することができた。

## (3) 時空間バースト検出手法

バーストとは事象が通常と比べて頻繁に現れている現象であり、バーストを検出することで当該事象の動向変化を明らかにすることができる。時間だけでなく、位置情報も考慮した時空間バースト検出手法を開発することで、時空間上で「どのように変化しているか」を分析可能となった。具体的には、Kelinbergの離散バースト検出モデルを拡張して、時空間的な変化でバーストを検出するための手法を開発した。図4に開発を行ったバースト検出手法で検出したバーストの検出結果を示す。図4では雪に関するリアルタイムに注目度の変化を捉えた様子を示している。

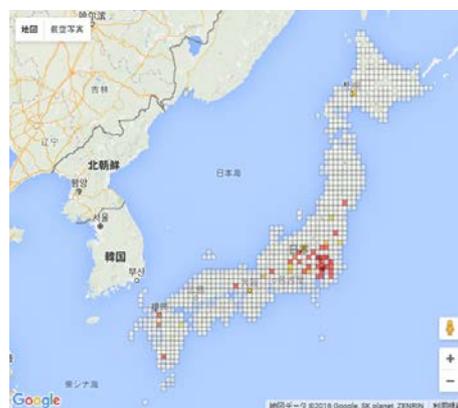


図4 バースト検出結果

## (4) 大規模ソーシャル画像データの効率的な管理法とその並列処理

大規模なソーシャル画像データをデータベース上で効率的に管理する方法を開発した。また、リアルタイムに動向分析が行えるようにするために、並列化による高速化を行った。

まず、内容ベースクラスタリング手法は、密度に基づくクラスタリング手法を用いてクラスタを取り出している。密度に基づくクラスタリング手法を高速化するために、並列化を行った。図5に示すように、空間的な領域をセルに分割しセル内のクラスタリングを並行して行う。セル間の統合処理に時間が必要となるため、均等な領域に分割するのではなく、複合グリッドと呼ばれる領域に分割することで効率的に並列化ができることを確認できた。

さらに、密度に基づくクラスタリング手法の新しいアルゴリズムとしてセルベースのクラスタリング手法を提案し、単一の時空間クラスタリングの高速化も行うことができた。

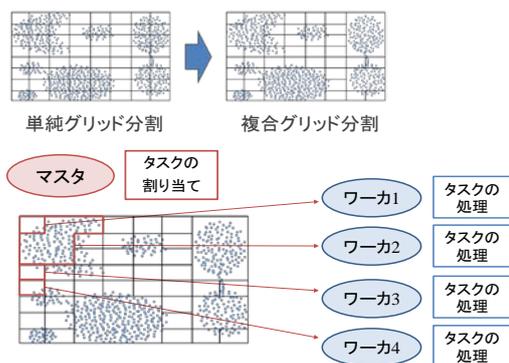


図5 クラスタリングの並列処理

また、バースト検出手法についても計算量が必要であり、データ分割並列化手法による並列化を行った。隠れマルコフモデルによって表現される Kleinberg のバースト検出手法について状態遷移コストの計算を並列に行うことで近似解を求める方法となっている。大規模なソーシャル画像データ集合に対して線形加速度に近い性能を持つ並列化手法であることを確認することができた。

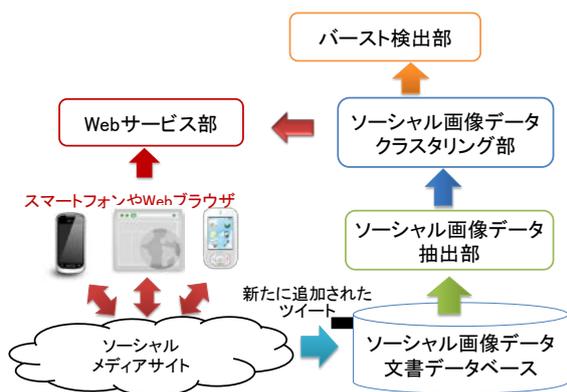


図6 作成したアプリケーション

### (5) 時空間上における動向情報の可視化

時空間的な変化をする動向情報について、画像データを含めて可視化する効率的な方法とユーザインタフェースを作成した。また、時空間クラスタとして取り出された話題が取り上げられている地域においてどのような話題が取り上げられているのか、代表ソーシャル画像を取り出す手法を開発した。図6に作成したWebアプリケーションの概要を示す。大雨を事例にソーシャル画像データを使用してその発生状況の変化を捉えることができるアプリケーションを実装し、リアルタイムに大雨の状況をモニタリングできることを確認することができた。

### 5. 主な発表論文等

[雑誌論文] (計 6 件)

- [1] Keiichi Tamura, Hajime Kitakami, and Tatsuhiko Sakai, Adaptive Distributed Modified Extremal Optimization for Maximizing Contact Map Overlap and Its Performance Evaluation, 査読有, International Journal Computational Intelligence Studies, (採録決定済).
- [2] Tatsuhiko Sakai, Keiichi Tamura, and Hajime Kitakami, Density-based Spatiotemporal Analysis System with Photo Image Classifier using the BoF Model, 査読有, Information Engineering Express (IEE), Vol.1, No.4, pp.85-94, 2015. <http://www.iaiai.org/journals/index.php/IEE/article/view/31/41>
- [3] 高橋誉文, 田村慶一, 黒木進, 北上始, 幾何学的接尾辞木の高速処理方式, 情報処理学会論文誌データベース, 査読有, Vol.8, No.2, pp.57-67, 2015年6月. <http://id.nii.ac.jp/1001/00142567/>
- [4] Shota Kotozaki, Keiichi Tamura, and Hajime Kitakami, Identifying local burstiness in a sequence of batched georeferenced documents, International Journal of Electronic Commerce Studies (IJECS), 査読有, Vol.6, No.2, pp.269-288, 2015. DOI: doi: 10.7903/ijecs.1347
- [5] Tatsuhiko Sakai and Keiichi Tamura, Real-time Analysis Application for Identifying Bursty Local Areas Related to Emergency Topics, 査読有, SpringerPlus, 4:162, April 2015.
- [6] Tatsuhiko Sakai, Keiichi Tamura, and Hajime Kitakami, Extracting Attractive Local-Area Topics in Georeferenced Documents using a New Density-based Spatial Clustering Algorithm, IAENG International Journal of Computer Science, 査読有, Volume 41 Issue 3, pp.131-140, August 2014. [http://www.iaeng.org/IJCS/issues\\_v41/issue\\_3/IJCS\\_41\\_3\\_04.pdf](http://www.iaeng.org/IJCS/issues_v41/issue_3/IJCS_41_3_04.pdf)

[学会発表] (計 39 件)

- [1] Tatsuhiko Sakai, Keiichi Tamura, and Hajime Kitakami, Cell-based DBSCAN Algorithm using Minimum Bounding Rectangle Criteria, BDMS 2017, March 27, 2017, 蘇州 (中国).
- [2] 酒井達弘, 田村慶一, 北上始, 密度に基づく時空間分析システムにおける学習済み深層ネットワークを用いた画像分類, インタラクティブ情報アクセスと可視化マイニング研究会(第15回), 2017年3月4日, 東京大学 (東京都).
- [3] Tatsuhiko Sakai, Keiichi Tamura, and Hajime Kitakami, Parallel Processing for Density-based Spatial Clustering Algorithm using Complex Grid Partitioning and Its

- Performance Evaluation, PDPTA 2016, July 25-28, 2016, ラスベガス (米国).
- [4] Keiichi Tamura, Tatsuhiro Sakai, and Hajime Kitakami, Location-based Temporal Burst Detection using Outlier Factors in Geo-tagged Tweets, 5th IIAI International Conference on Advanced Applied Informatics IIAI-AAI 2016, July 10-14, 2016, 熊本市国際交流会館 (熊本市).
- [5] Tatsuhiro Sakai, Keiichi Tamura, and Hajime Kitakami, Real-time Local Topic Extraction using Density-based Adaptive Spatiotemporal Clustering for Enhancing Local Situation Awareness, KDIR2017, November 12-14, 2015, リスボン (ポルトガル).
- [6] 酒井 達弘, 田村 慶一, 事崎翔太, 林田翼沙, 北上始, 密度に基づく時空間分析手法を用いた実世界トピックの閲覧システム, 第17回 IEEE 広島支部学生シンポジウム, 2015年11月19日, 20日, 岡山大学 (岡山市).
- [7] 事崎 翔太, 田村 慶一, 北上始, バースト度の地域的な変化の可視化に関する検討, 第17回 IEEE 広島支部学生シンポジウム, 2015年11月19日, 20日, 岡山大学 (岡山市).
- [8] Tatsuhiro Sakai, Keiichi Tamura, and Hajime Kitakami, Identifying Main Topics in Density-based Spatial Clusters using Network-based Representative Document Extraction, IEEE IWCI2015, November 6-7, 2015, 広島市立大学 (広島市).
- [9] Keiichi Tamura, Tomoki Matsui, Hajime Kitakami, and Tatsuhiro Sakai, Identifying Local Temporal Burstiness using MACD Histogram, IEEE SMC2015, October 9-12, 2015, 香港 (中国).
- [10] Tatsuhiro Sakai, Keiichi Tamura, and Hajime Kitakami, Extracting Topic-related Photos in Density-based Spatiotemporal Analysis System for Enhancing Situation Awareness, 4th IIAI International Conference on Advanced Applied Informatics IIAI-AAI 2015, July 12-16, 2015, 岡山コンベンションセンター (岡山市).
- [11] Tatsuhiro Sakai, Keiichi Tamura, and Hajime Kitakami, Emergency Situation Awareness during Natural Disasters using Density-based Adaptive Spatiotemporal Clustering, BDMS2015, April 20, 2015, ハノイ (ベトナム).
- [12] Keiichi Tamura, and Tatsuhiro Sakai, Density-based Semantic Spatial Clustering for Extracting Areas of Interesting in Geo-tagged Photo Images, CATA-2015, March 9-11, 2015, ホノルル (米国).
- [13] Tatsuhiro Sakai, and Keiichi Tamura, Identifying Bursty Areas of Emergency

Topics in Geotagged Tweets using Density-based Spatiotemporal Clustering Algorithm, IWCI2014, November 7-8, 2014, 広島市立大学 (広島市).

- [14] Keiichi Tamura and Hajime Kitakami, A New Parallelization Model for Detecting Temporal Bursts in Large-Scale Document Streams on a Multi-Core CPU, SMC2014, October 5-8, 2014, サンディエゴ (米国).
- [15] Tatsuhiro Sakai, Keiichi Tamura, and Hajime Kitakami, Density-based Adaptive Spatial Clustering Algorithm for Identifying Local High-Density Areas in Georeferenced Documents, SMC2014, October 5-8, 2014, サンディエゴ (米国).
- [16] Shota Kotozaki, Keiichi Tamura, and Hajime Kitakami, A New Method for Identifying Location-Based Bursts in a Sequence of Batched Georeferenced Documents, Nets2014, August 16-17, 2014, シンガポール.

[図書] (計0件)

[産業財産権]

○出願状況 (計0件)

○取得状況 (計0件)

[その他]

なし

## 6. 研究組織

### (1) 研究代表者

田村 慶一 (TAMURA, Keiichi)

広島市立大学・情報科学研究科・准教授

研究者番号: 80347616