

令和 3 年 6 月 8 日現在

機関番号：25403

研究種目：基盤研究(C)（一般）

研究期間：2018～2020

課題番号：18K11320

研究課題名（和文）実世界情報分析のための時空間データマイニングに関する研究

研究課題名（英文）Spatio-Temporal Data Mining for Real World Information Analysis

研究代表者

田村 慶一（Tamura, Keiichi）

広島市立大学・情報科学研究科・教授

研究者番号：80347616

交付決定額（研究期間全体）：（直接経費） 3,400,000円

研究成果の概要（和文）：近年、実世界で人々が目にした出来事や耳にした生の声がソーシャルメディア上で時間と位置情報を持つ時空間データとして投稿され、重要な情報源となっている。ソーシャルメディア上で生成される時間、位置と内容（コンテンツ）を持つデータを時空間ソーシャルデータと呼ぶ。時空間ソーシャルデータを対象とした時空間データマイニング技術を開発し、「何が、いつ、どこで発生し、どのように変化しているか」という情報を抽出することが可能となった。

研究成果の学術的意義や社会的意義

時空間ソーシャルデータを対象とした時空間データマイニング技術を用いることでソーシャルメディア上に投稿されている情報をリアルタイムに把握することができ、観光情報、地域振興、マーケティング、防災や危機管理の情報源として時空間ソーシャルデータを有効活用することが可能となる。また、実世界の事象を多面的に分析可能となり、ソーシャルメディアのICTへの利活用に新しいイノベーションをもたらすことができる。

研究成果の概要（英文）：In these days, people post geo-social data with time and location information on social media. These posts include things that people are witnessing and they are related to real comments in the real world. Geo-spatial data including time, location and content is called geo-spatial social data. In this study, we have developed spatio-temporal data mining techniques for geo-spatial social data. These spatio-temporal data mining techniques enable us to know what is happening, when the things are happening, where the things are happening, how things are changing. We can use geo-spatial social data as an information source by using spatio-temporal data mining techniques for geo-spatial social data.

研究分野：データ工学

キーワード：時空間データマイニング ソーシャルメディア マルチモーダル 深層学習 高性能データマイニング

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

### 1. 研究開始当初の背景

GPS 付きスマートフォンの普及と地理情報サービスと連携したアプリケーションの利用拡大とともに、実世界で人々が目にした出来事や耳にした生の声が、時間と位置情報を持つ時空間データとしてソーシャルメディア上に盛んに投稿されている。人や物をセンサととらえると、投稿されるデータはセンサが測定した実世界の観測データとして扱うことができ、観光情報、地域振興、マーケティング、防災や危機管理における重要な情報源として期待されている。しかしながら、基礎研究や事例研究が盛んに行われる一方、投稿内容が十分に活用されない事例や、その利活用も、新聞、テレビや Web 上での情報発信を補完する程度に留まっている。利活用が限定的になっている原因として、複数メディアを統合する技術や、また、情報を集約するための時空間データマイニング技術が不十分であることがあげられる。

### 2. 研究の目的

ソーシャルメディア上において投稿される時間、位置と内容(コンテンツ)を持つデータを時空間ソーシャルデータと呼ぶ。コンテンツとして、テキスト、画像、動画など多種多様なデータが日々投稿され、投稿されるデータは実世界の事象と結びついた時空間集合知をサイバー空間に形成しつつある(図1)。本研究では、時空間ソーシャルデータを用いた実世界情報分析を対象とし、新しい時空間データマイニングの基盤技術の確立を目指す。具体的には、複数のメディアを統合的に扱うために、複数のメディア(モダリティ)間の関係を定義可能で、また、マルチモーダルデータを同一空間で扱うためのマルチモーダル特徴量空間(図1)、「何が、いつ、どこで発生し、どのように変化しているか」という事象の時空間的な推移を集約する技術、また、大規模データを対象とした高性能マイニング手法と可視化技術の開発を行う。

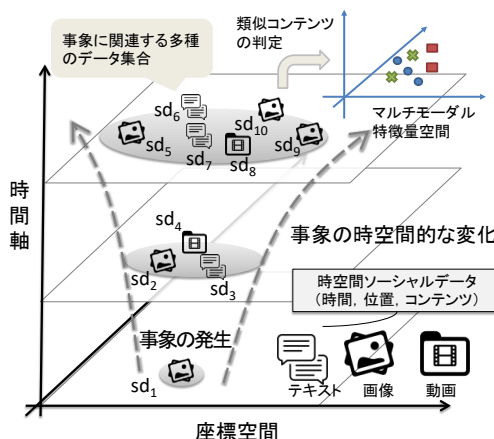


図1 時空間ソーシャルデータ

### 3. 研究の方法

本研究では、実世界情報分析を対象に、次の5つの研究サブテーマに取り組み、時空間ソーシャルデータを用いた新しい時空間データマイニング技術の開発を行った。

- (1) マルチモーダル特徴量空間
- (2) 動的な時空間クラスタリング手法
- (3) オートキュレーション技術
- (4) 大規模時空間ソーシャルデータの効率的な管理
- (5) 時空間ソーシャルデータの情報可視化

### 4. 研究成果

#### (1) マルチモーダル特徴量空間

深層学習を利用し、多種多様なデータを横断的に扱うことができるマルチモーダル特徴量空間の開発を行った。深層学習において中間層に現れる特徴量をボトルネック特徴量と呼ぶ。ここでは、複数のコンテンツが添付されているデータを利用して、マルチモーダルデータから取り出したボトルネック特徴量をひとつの統合空間に写像する方式を開発した。テキストデータを対象にテキスト本文、また、テキスト本文が示す品詞情報を1つの空間(マルチチャンネル空間)にまとめて当該テキストデータの特徴量を抽出することでテキスト分類精度を向上させることができることを確認した(図2)。また、特徴量を取り出すための深層学習のモデルの改良を行い、ベンチマークデータセットを用いた評価実験での検証を行った。

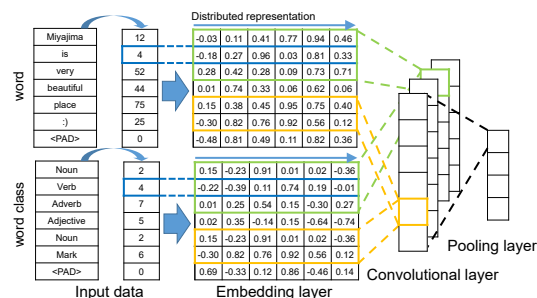


図2 マルチチャンネルを利用した特徴空間

## (2) 動的な時空間クラスタリング手法

時空間ソーシャルデータから事象を抽出するための時空間クラスタリング手法を作成した。また、時空間クラスタの時空間的な変化を追跡可能とするために、時空間クラスタの状態遷移を検出可能なアルゴリズムを設計した。具体的には、バースト度によって重要度を算出する方法、AnyTime アルゴリズムを応用して段階的にクラスタリング結果を取り出すアルゴリズムを開発した。

## (3) オートキュレーション技術

抽出された時空間クラスタの中身がどのようなものであるかを自動的に抽出する必要がある。そこで、ネットワークグラフベースの要約手法を応用し、時空間的な要約を作成可能とする手法を作成した。データの類似度によってグラフ構造を作成し、グラフ上の重要ノードを見つけ、重要度によってランキングを行う。ランキング上位のデータ内容を提示することで時空間クラスタの要約とする方法や重要キーワードを提示する方法を開発した。

## (4) 大規模時空間ソーシャルデータの効率的な管理

大規模な時空間ソーシャルデータを効率的に管理する方法と、リアルタイムに処理を実行するために、並列化による高速化手法の開発を行った。時空間ソーシャルデータの効率的な管理とその並列処理に関しては、基盤となる時空間クラスタリングの高速化と、並列化可能な処理モデルを作成した (図3と図4)。また、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、順序付けを行う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を行うことができた。

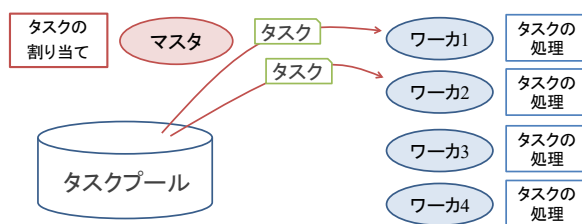


図3 空間クラスタリングの並列化

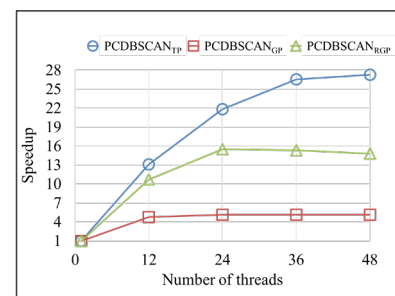


図4 並列化性能

## (5) 時空間ソーシャルデータの情報可視化

時空間ソーシャルデータから抽出された各イベントやトピックを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、また、いつどこで話題が取り上げられているかを分かりやすくヒートマップなどで表示するためのシステムを開発した。具体的には、観光情報 (図5) と災害情報の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、リアルタイムに状況をモニタリングできることを確認できた。

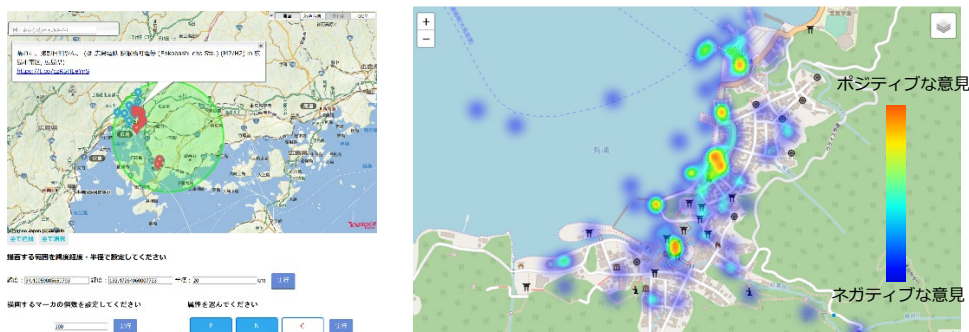


図5 時空間ソーシャルデータの可視化

空間ソーシャルデータを対象とした新しい時空間データマイニング技術によって、時空間ソーシャルデータを統合的に分析可能となり、今後、これらの技術を用いて、ソーシャルメディア上に投稿されるデータのICTへの利活用への応用が期待できる。

## 5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 4件／うち国際共著 0件／うちオープンアクセス 0件）

1. 著者名 Keiichi Tamura, Akitada Omagari, Hajime Ito, Shuichi Hashida	4. 巻 No. 2/3
2. 論文標題 Detecting Audio Adversarial Examples for Protecting Speech-to-Text Transcription Neural Network	5. 発行年 2021年
3. 雑誌名 International Journal of Computational Intelligence Studies	6. 最初と最後の頁 採択決定済み
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 橋田 修一, 田村 慶一	4. 巻 Vol.13, No.2
2. 論文標題 Multi-Channel MHLFを用いた時系列データの分類手法	5. 発行年 2020年
3. 雑誌名 情報処理学会論文誌 数理モデルとその応用(TOM)	6. 最初と最後の頁 22 ~ 35
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Keiichi Tamura; Takumi Ichimura	4. 巻 Vol.7 No.3/4
2. 論文標題 Time series classification using MACD-histogram-based recurrence plot	5. 発行年 2018年
3. 雑誌名 International Journal of Computational Intelligence Studies	6. 最初と最後の頁 192 ~ 213
掲載論文のDOI（デジタルオブジェクト識別子） 10.1504/IJCISTUDIES.2018.096188	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Shuichi Hashida, Keiichi Tamura, Tatsuhiko Sakai	4. 巻 Vol. 46, No. 1
2. 論文標題 Classifying Tweets using Convolutional Neural Networks with Multi-Channel Distributed Representation	5. 発行年 2019年
3. 雑誌名 IAENG International Journal of Computer Science	6. 最初と最後の頁 68 ~ 75
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計22件（うち招待講演 0件 / うち国際学会 6件）

1. 発表者名 Jun Teraoka and Keiichi Tamura
2. 発表標題 Detecting Adversarial Examples for Time Series Classification and its Performance Evaluation
3. 学会等名 KES-IDT2021(採択決定済) (国際学会)
4. 発表年 2021年

1. 発表者名 Keiichi Tamura and Shuichi Hashida
2. 発表標題 (n, m)-Layer MC-MHLF: Deep Neural Network for Classifying Time Series
3. 学会等名 2020 IEEE International Conference on Systems, Man, and Cybernetics (国際学会)
4. 発表年 2020年

1. 発表者名 山本 愛海, 田村 慶一
2. 発表標題 深層学習を用いたマルチラベル分類に基づく災害画像分類
3. 学会等名 電子情報通信学会2021年総合大会
4. 発表年 2021年

1. 発表者名 玄行 朱里, 田村 慶一
2. 発表標題 蒸留を用いた時系列分類モデルMC-MHLFの圧縮
3. 学会等名 情報処理学会第83回全国大会
4. 発表年 2021年

1. 発表者名 寺岡 純, 田村 慶一
2. 発表標題 時系列データ分類問題に対する敵対的サンプルの検出手法
3. 学会等名 2020 IEEE SMC Hiroshima Chapter 若手研究会
4. 発表年 2020年

1. 発表者名 Keiichi Tamura, Akitada Omagari, and Shuichi Hashida
2. 発表標題 Novel Defense Method against Audio Adversarial Example for Speech-to-Text Transcription Neural Networks
3. 学会等名 2019 IEEE 11th International Workshop on Computational Intelligence and Applications (IWCIA) (国際学会)
4. 発表年 2019年

1. 発表者名 Shuichi Hashida, Keiichi Tamura
2. 発表標題 Multi-Channel MHLF: LSTM-FCN using MACD-Histogram with Multi-Channel Input for Time Series Classification
3. 学会等名 2019 IEEE 11th International Workshop on Computational Intelligence and Applications (IWCIA) (国際学会)
4. 発表年 2019年

1. 発表者名 Shuichi Hashida, Keiichi Tamura
2. 発表標題 MACD-Histogram-based Fully Convolutional Neural Networks for Classifying Time Series
3. 学会等名 The 2019 6th International Conference on Control, Decision and Information Technologies (CODIT'19) (国際学会)
4. 発表年 2019年

1. 発表者名 三木 直人, 酒井達弘, 田村 慶一
2. 発表標題 セル結合判定の順序付けに基づくCDBSCANアルゴリズムの高速化
3. 学会等名 電子情報通信学会2020年総合大会
4. 発表年 2020年

1. 発表者名 橋田 修一, 田村 慶一
2. 発表標題 Multi-Channel MHLFを用いた時系列データの分類手法
3. 学会等名 第126回数理解モデル化と問題解決研究発表会
4. 発表年 2019年

1. 発表者名 橋田 修一, 田村 慶一
2. 発表標題 MACDヒストグラムを用いた深層学習による時系列データ分類手法
3. 学会等名 WebDB Forum 2019
4. 発表年 2019年

1. 発表者名 甲斐 健太, 田村 慶一, 橋田 修一
2. 発表標題 超高密度気象観測データを用いた深層学習による暑さ指数の予測
3. 学会等名 2019 IEEE SMC Hiroshima Chapter 若手研究会
4. 発表年 2019年

1. 発表者名 橋田 修一, 田村 慶一
2. 発表標題 MACDヒストグラムを用いたMulti-Channel MHLFによる時系列データ分類手法
3. 学会等名 2019 IEEE SMC Hiroshima Chapter若手研究会
4. 発表年 2019年

1. 発表者名 橋田修一, 田村 慶一, 酒井達弘
2. 発表標題 部分系列のクラスタリングに基づく符号化を用いたCNNによる時系列データの分類手法
3. 学会等名 2018 IEEE SMC Hiroshima Chapter若手研究会
4. 発表年 2018年

1. 発表者名 Shuichi Hashida, Keiichi Tamura, Tatsuhiko Sakai
2. 発表標題 Classifying Sightseeing Tweets using Convolutional Neural Networks with Multi-Channel Distributed Representation
3. 学会等名 2018 IEEE International Conference on Systems, Man, and Cybernetics (国際学会)
4. 発表年 2018年

1. 発表者名 村上和希, 橋田修一, 田村 慶一, 酒井達弘
2. 発表標題 深層学習による分類に基づく観光ツイートの分析手法
3. 学会等名 平成30年度(第69回)電気・情報関連学会中国支部連合大会
4. 発表年 2018年



1. 発表者名 橋田修一, 田村 慶一
2. 発表標題 クラスタリングに基づく符号化手法を用いたCNNによる時系列データの分類
3. 学会等名 平成30年度(第69回)電気・情報関連学会中国支部連合大会
4. 発表年 2018年

1. 発表者名 橋田修一, 田村 慶一
2. 発表標題 CNNによる時系列データ分類のための符号化手法とその評価
3. 学会等名 測自動制御学会 システム・情報部門学術講演会2018
4. 発表年 2018年

1. 発表者名 酒井達弘, 田村 慶一, 北上始, 竹澤寿幸
2. 発表標題 セルベースのDBSCANのAnytimeアルゴリズム
3. 学会等名 第11回データ工学と情報マネジメントに関するフォーラム (DEIM2019)
4. 発表年 2019年

1. 発表者名 尾曲晃忠, 橋田修一, 田村 慶一
2. 発表標題 Audio Adversarial Examplesに対する動的リサンプリング法とノイズ除去法による防御
3. 学会等名 情報処理学会第81回全国大会
4. 発表年 2019年

1. 発表者名 橋田修一, 田村 慶一
2. 発表標題 ドアの開閉動作を用いたShapeletによる個人識別手法
3. 学会等名 電子情報通信学会2019年総合大会
4. 発表年 2019年

1. 発表者名 甲斐健太, 橋田修一, 田村 慶一
2. 発表標題 超高密度気象観測データを用いたLSTMによる暑さ指数の予測
3. 学会等名 電子情報通信学会2019年総合大会
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関