

科学研究費助成事業 研究成果報告書

令和 3 年 6 月 7 日現在

機関番号：14301

研究種目：基盤研究(B)（一般）

研究期間：2018～2020

課題番号：18H03245

研究課題名（和文）Web情報に基づく社会情報分析と行動予測技術の開発

研究課題名（英文）Development of Methods for Social Information Analysis and Behavior Prediction Based on Web Information

研究代表者

田島 敬史（Tajima, Keishi）

京都大学・情報学研究科・教授

研究者番号：60283876

交付決定額（研究期間全体）：（直接経費） 13,300,000円

研究成果の概要（和文）：Web上から現実社会の事象に関する情報を収集するための技術として、ある語がWeb検索される頻度とWeb上に現れる頻度の比較から現実社会の事象に関係する語を識別する手法や、ソーシャルメディア上に流れる情報のうち、現在起こっている事象に関するものとそうでないものを区別する手法などを開発した。

また、Webなどから抽出した時間、地域、キーワードなどの複数の属性からなる大規模な時系列データから、様々なイベント発生の兆候と、それらとの因果関係をリアルタイムに抽出し、将来予測を高速かつ完全自動に、また、状況変化に対応しながら長期にわたって、行う新技術を開発した。

研究成果の学術的意義や社会的意義

Web上の情報から実世界の状況を素早く認識できるようになることは、市場トレンドや行政ニーズの把握、災害時の迅速な救助、犯罪等の緊急情報の検出などを可能にし、社会的意義は大きい。また、そのような分析を行う場合、根拠となる情報の抽出が不正確あるいは偏っていると、分析の結果も不正確あるいは偏ったものとなるため、正確かつ多様な情報を抽出する技術は社会的に重要である。さらに、事象間の関連性のモデル化は、予測のためだけでなく、社会現象や人間の行動のより深い理解にも有用である。また、技術的には、大衆活動と個人活動との関連性を分類しながら学習する点や、与えられた時間内にリアルタイム処理を行う点も特徴である。

研究成果の概要（英文）：We developed several techniques for collecting information related to real-world events from the Web, such as a technique for identifying keywords related to real-world events by comparing the frequency of Web search by the words and the frequency of their appearance on the Web, and a technique for distinguish information on social media related to real-world events.

We also developed techniques for monitoring time-evolving data consisting of time, place, keywords, and so on extracted from the Web, identifying signals of new events and causal relationship between them in real-time, and predicting the future events efficiently and fully automatically for long time with adapting the change of the trend.

研究分野：Web情報分析，ソーシャルネットワーク分析，情報検索，データベースシステム

キーワード：オンライン・アクティビティ抽出 ソーシャルメディア分析 時系列データ 情報有効期限 データストリーム予測 要因分析 特徴抽出 相互作用発見

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

本研究開始当時までにWebには一般ユーザからの情報が大量に蓄積されるようになっており、特にソーシャルメディアの普及により、リアルタイムの情報もWebから豊富に得られるようになっていた。その結果、Webを社会の動向をリアルタイムに反映するセンサーとみなし、Web上の情報からオフライン社会に関する情報を抽出する研究が盛んに行われていた。

2. 研究の目的

人々の行動や社会現象の発生と終了には要因となる別の事象があり、事象間には相互の関連性がある。そして、Web情報はこれらの事象の時間変化と事象間の関連性についても有用な情報源となっていると考えられる。そこで、本研究では、これらの事象の変化と連鎖の情報をWebから抽出する技術と、これらの変化や連鎖を表現できるモデルを開発し、また、このモデルに基づいて社会情報分析と行動予測を行う技術を開発することを目的とした。

3. 研究の方法

Webからの情報抽出については、研究代表者である田島がこれまで開発してきたWeb検索技術およびソーシャルネットワーク分析技術をさらに発展させ、また、これらを基盤としてさらなる技術を開発する。また、事象間の連鎖のモデルと予測については、研究分担者である櫻井がこれまで開発してきたテンソル解析技術を深化させることで実現する。

4. 研究成果

Webからの情報抽出については、以下の技術を開発した。

検索エンジンへの入力を利用した現実世界の事象に関連する語の検出技術

本研究で、過去の様々な現実社会のイベントに関連する語のWeb上の出現を分析した結果、ある語を含む新規Webページの出現頻度は大きく変化していないにも関わらず、その語が検索エンジンへ入力される頻度のみが急上昇した場合には、なんらかの現実世界の事象が関係している確率が高いことを発見し、この結果に基づき、検索エンジンへ入力される検索語の頻度の変化と、それらの語を含む新規Webページの出現頻度の変化の情報を用いて、Web上の大量の情報の中から現実社会の現在の事象に関連するものを発見するための手法を開発した。

ソーシャルメディアの情報の現在の事象に関連する情報と関連しない情報への分類技術

Twitterなどのソーシャルメディア上には、特定の現在の事象には関係しない時間依存性が低い投稿と、特定の現在の事象に関係していて時間依存性が高い投稿の双方が流れる。様々な活動の時間変化を知るためにソーシャルメディアを用いるためには、これら二種類を区別する必要がある。時間依存性が高い投稿の情報においては、その価値に有効期限があり、その期限を過ぎると急激に情報の価値が低下することが多い。この点に着目して、まず、リツイートなどの情報に対する反応の時間的変化にもとづいて情報の有効期限を推定する手法を開発した。

しかし、これだけでは、ある情報の有効期限が実際に来るまで、その情報が有効期限がある時間依存性の高い情報なのか、有効期限がない時間依存性の低い情報なのかはわからない。そこで、この判定手法を用いて過去のソーシャルメディア上の大量の投稿について有効期限を判定し、これらのデータを機械学習により学習させることで、ソーシャルメディア上の情報のテキストと、それに対するリツイートなどのこれまでの反応の情報から、将来の有効期限を推定するモデルを学習する手法を開発した。

ソーシャルメディアの情報からの重要なキーワードの検出技術

Twitterなどのソーシャルメディアの非常に短い文章から重要な事象に関連する語を識別する際に必要となる要素技術として、各投稿中で重要な役割を果たしている語を識別する技術を開発した。開発した手法では、投稿内で重要な役割を果たしている語は、その語が表すトピックと投稿全体のトピックとの類似度が高いという仮定に基づき、各語と投稿全体のそれぞれをトピック空間中のベクトルへと変換して、このベクトル間の類似度を用いて

重要な語を判定する。

ある語に関する多様なトピックの情報を収集する技術

現実世界の現在の事象に関連すると思われる語を検出した場合、その語を含む Web 上の情報を収集することによって、その現在の事象に関する情報を収集することを考えたい。しかし、その語に関連する、より代表的な大きなトピックがある場合、その語による通常の Web 上の様々な情報の検索では、その大きなトピックに関する情報で検索結果が占められてしまい、他の話題に関する情報が収集できないことがある。そこで、小さな話題の各々に関連するフレーズを大量に自動生成し、これらを用いたフレーズ検索を行うことにより、多様な情報を収集する技術を開発した。

また、抽出した時系列データから事象間の関係のモデル化や今後の事象の予測を行う手法に関しては、研究分担者である櫻井がこれまで開発したリアルタイム AI 技術をさらに発展させ、以下のような特徴抽出技術、要因分析・予測技術を開発した。

自動パターン検出のためのストリームアルゴリズム：StreamScope

StreamScope は大規模時系列データストリームに含まれる典型的なパターンを発見するための手法であり、与えられたデータストリームをリアルタイムに解析し、時系列パターンの種類と変化点を発見し、それらの特徴をモデルパラメータとして表現する。本手法では符号化理論に基づくモデル評価基準を応用し、解析データに関する事前知識を必要とせず、データの要約情報（時系列パターンの種類・変化点）を自動的に抽出する。図 1 は本手法を用いたリアルタイム解析の出力例である。図上段はヒトの両手足 4 箇所に加速度センサを設置し、歩行と 3 種類（両腕、右腕、左腕）のストレッチで構成される合計 4 種類の動作を捉えたデータストリームを示す。図下段の番号は、提案手法が検出したパターンの ID を示し、長方形の両端は各パターンの開始点・終了点を示す。提案手法は 8 つのパターン変化点を自動検出し、それらを 4 つの動作へと分類することに成功している。本手法は Web 情報のオンライン解析にも有用であり、論文は Web 情報検索分野のトップ国際会議の一つである CIKM に採択されている。

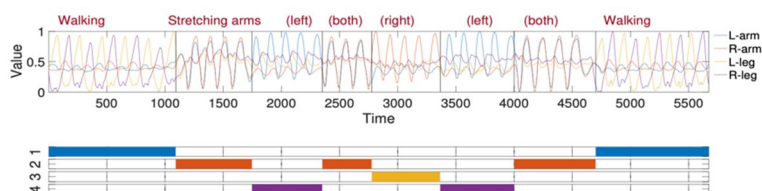


図 1 モーションキャプチャデータストリームに対する StreamScope の出力結果

リアルタイム要因分析技術：OrbitMap

OrbitMap は複合ビッグデータのための要因分析・予測手法である。Web データストリームをはじめとする大規模な時系列データに対し、リアルタイムに重要な特徴や潜在的なトレンド（レジーム）を発見し、各レジーム間の動的な関係性を抽出することで、長期的かつ継続的に時系列イベントストリーム内の重要な動的要因を監視し、将来のイベント予測を行う。図 2 は提案手法における動的空間推移モデルの概要を示している。図 2(a) は単一のレジームにおける潜在空間内での動的パターン生成の様子、図 2(b) は、2 つの異なるレジーム間での動的空間遷移の様子、そして、図 2(c) は複数（ここでは、 $r = 5$ ）のレジーム間での動的空間遷移および、それに伴う動的パターン生成の様子を示している。提案モデルは既存の動的モデルと異なり、複数の異なる非線形モデル間を遷移することで、複雑な時系列パターンを表現する。

そして提案手法においては、時系列イベントストリームからリアルタイム要因分析と将来イベント予測を行う。リアルタイムに現在の時系列パターン（レジーム）とその変化点を発見し、それらを動的空間遷移ネットワーク M に格納する。過去に同様のレジームが出現した場合には、そのネットワークをたどることで、一定時刻後の未知のイベントを予測する。論文はデータマイニング分野のトップ国際会議である KDD2019 に採択されている。

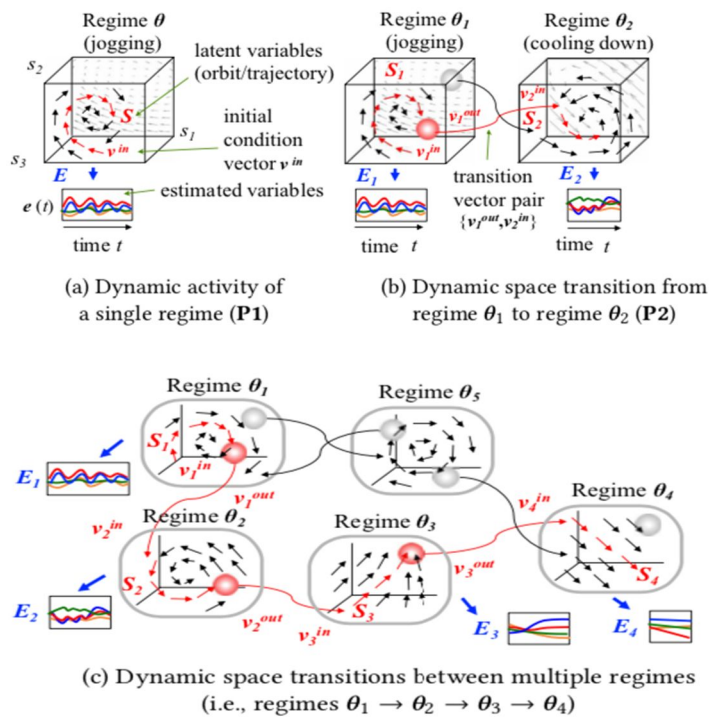


図2 提案モデルにおける動的空間推移ネットワークの様子

オンライン活動データのリアルタイム予測技術：CubeCast

CubeCast は大規模オンライン活動データのための特徴自動抽出・リアルタイム予測手法である。図3に提案手法の概要を示す。本手法は、時間、地域、キーワードのように複数の属性を持つテンソルデータストリームが与えられたとき、最新の観測データ（図3(a)青）を監視しながら潜在的なトレンドを発見し、適応的にモデルを変化させながら長期先のデータ（図3(a)赤）を予測し続ける。このとき、図3(b)のように各地域で共通する季節パターンを抽出し、それらに基づき図3(c)のように類似パターンを有する地域のグループ化を行う。また、提案モデルは非線形性を有する長期トレンド、季節トレンドを同一のモデル空間で表現し、それらの相互作用を抽出する。例えば、ある商品のWeb検索数が年末セールの際に増加する傾向を持つとき、その増加率は平常時の検索数の増加に伴い相乗的に大きくなることもある。このような現象をモデル化することにより、予測精度の向上を実現した。論文はデータマイニング分野のトップ国際会議であるKDD2020に採択されている。

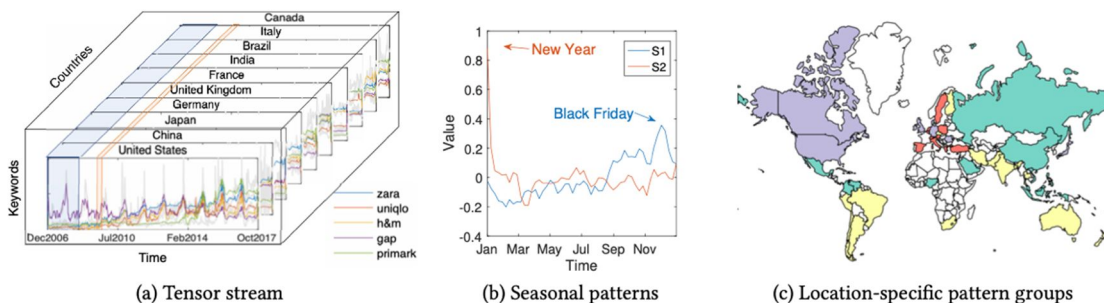


図3 提案手法によるリアルタイム予測の様子と検出パターンの例

5. 主な発表論文等

〔雑誌論文〕 計8件（うち査読付論文 8件/うち国際共著 0件/うちオープンアクセス 3件）

1. 著者名 Masahiro Inoue, Keishi Tajima	4. 巻 -
2. 論文標題 Temporal Analysis of Supply and Demand of Topics on The Web	5. 発行年 2019年
3. 雑誌名 Proceedings of 11th ACM Conference on Web Science	6. 最初と最後の頁 143-144
掲載論文のDOI（デジタルオブジェクト識別子） 10.1145/3292522.3326054	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Yasuko Matsubara, Yasushi Sakurai	4. 巻 -
2. 論文標題 Dynamic Modeling and Forecasting of Time-evolving Data Streams	5. 発行年 2019年
3. 雑誌名 Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)	6. 最初と最後の頁 印刷中
掲載論文のDOI（デジタルオブジェクト識別子） 10.1145/3292500.3330947	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Kouki Kawabata, Yasuko Matsubara, Yasushi Sakurai	4. 巻 -
2. 論文標題 StreamScope: Automatic Pattern Discovery over Data Streams	5. 発行年 2018年
3. 雑誌名 Proceedings of ACM SIGMOD Workshop on Exploiting Artificial Intelligence Techniques for Data Management (aiDM)	6. 最初と最後の頁 5:1-5:8
掲載論文のDOI（デジタルオブジェクト識別子） 10.1145/3211954.3211959	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 川畑光希, 松原靖子, 櫻井保志	4. 巻 Vol.11, No.1
2. 論文標題 自動パターン検出のためのストリームアルゴリズム	5. 発行年 2018年
3. 雑誌名 情報処理学会論文誌:データベース	6. 最初と最後の頁 1-10
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 Koki Kawabata, Yasuko Matsubara, Takato Honda, Yasushi Sakurai	4. 巻 -
2. 論文標題 Non-Linear Mining of Social Activities in Tensor Streams	5. 発行年 2020年
3. 雑誌名 Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)	6. 最初と最後の頁 2093-2102
掲載論文のDOI (デジタルオブジェクト識別子) 10.1145/3394486.3403260	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Thinh Minh Do, Yasuko Matsubara, Yasushi Sakurai	4. 巻 Vol. 28
2. 論文標題 Real-time Forecasting of Non-linear Competing Online Activities	5. 発行年 2020年
3. 雑誌名 Journal of Information Processing	6. 最初と最後の頁 333-342
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 木村 輔, 松原 靖子, 川畑 光希, 櫻井 保志	4. 巻 Vol. 14, No. 2
2. 論文標題 大規模疫病データの ための将来予測アルゴリズム	5. 発行年 2021年
3. 雑誌名 情報処理学会論文誌: データベース	6. 最初と最後の頁 10-19
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 川畑 光希, 松原 靖子, 本田 崇人, 櫻井 保志	4. 巻 Vol. 14, No. 3
2. 論文標題 オンライン活動データ ストリームのための非線形モデル解析	5. 発行年 2021年
3. 雑誌名 情報処理学会論文誌: データベース	6. 最初と最後の頁 -
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15件（うち招待講演 11件 / うち国際学会 1件）

1. 発表者名 長島弘昂, 田島敬史
2. 発表標題 周辺語に基づく有効期限を表す時間表現の判定
3. 学会等名 第13回データ工学と情報マネジメントに関するフォーラム
4. 発表年 2021年

1. 発表者名 山田瑛平, 田島敬史
2. 発表標題 SNSにおけるグラフの時間変化を考慮したユーザ推薦手法
3. 学会等名 第13回データ工学と情報マネジメントに関するフォーラム
4. 発表年 2021年

1. 発表者名 川畑 光希, 松原 靖子, 本田 崇人, 今井 優作, 田嶋 優樹, 櫻井 保志
2. 発表標題 大規模購買ログの時系列分析に基づくLTV予測
3. 学会等名 2020年度人工知能学会全国大会（第34回）論文集, 4K2-GS-3-03
4. 発表年 2020年

1. 発表者名 櫻井保志
2. 発表標題 IoTビッグデータのためのリアルタイムAI技術とその応用
3. 学会等名 組込みシステム産業振興機構技術セミナー（招待講演）
4. 発表年 2020年

1. 発表者名 櫻井保志
2. 発表標題 スマート工場のための人工知能技術基盤～ビッグデータ解析に基づくリアルタイムAIや知的支援システムの開発
3. 学会等名 NEDO「IoT推進のための横断技術開発プロジェクト」スクール・ワークショップ（招待講演）
4. 発表年 2020年

1. 発表者名 櫻井保志
2. 発表標題 ビッグデータのためのリアルタイムAI技術
3. 学会等名 テクノアリーナフォーラム「先読みシミュレーション」+物質・材料科学研究推進機構講演会（招待講演）
4. 発表年 2020年

1. 発表者名 Yasushi Sakurai
2. 発表標題 Invited Talk: Real-time AI technologies for Big Data: Foundations and Challenges
3. 学会等名 Osaka University / University of Groningen Workshop（招待講演）（国際学会）
4. 発表年 2020年

1. 発表者名 櫻井保志
2. 発表標題 時系列ビッグデータのためのリアルタイムAI技術
3. 学会等名 一般社団法人日本電気計測器工業会（JEMIMA）先端技術調査委員会技術講演会（招待講演）
4. 発表年 2020年

1. 発表者名 櫻井保志
2. 発表標題 IoTビッグデータのためのリアルタイムAI技術
3. 学会等名 プリンテッド・エレクトロニクス研究会2019年度第4回PE研究会（招待講演）
4. 発表年 2020年

1. 発表者名 櫻井保志
2. 発表標題 IoTデータのためのリアルタイムAI技術と製造業の高度化
3. 学会等名 一般社団法人 京都知恵産業創造の森スマート社会推進フォーラム（招待講演）
4. 発表年 2020年

1. 発表者名 Yasushi Sakurai
2. 発表標題 Real-time Forecasting of IoT Big Data: Foundations and Challenges
3. 学会等名 AIRC-ISIR International Symposium（招待講演）
4. 発表年 2019年

1. 発表者名 櫻井保志
2. 発表標題 IoTビッグデータのためのAI技術とその応用
3. 学会等名 電気材料技術懇談会・特別講演会（招待講演）
4. 発表年 2019年

1. 発表者名 櫻井保志
2. 発表標題 IoTビッグデータのリアルタイム解析とその応用
3. 学会等名 人工知能学会計測インフォマティクス研究会（招待講演）
4. 発表年 2019年

1. 発表者名 櫻井保志
2. 発表標題 IoTビッグデータのためのAI技術とその応用
3. 学会等名 一般社団法人電子情報技術産業協会（JEITA）2019技術セミナー（招待講演）
4. 発表年 2019年

1. 発表者名 岩永雅史，田島敬史，山肩洋子
2. 発表標題 Twitterにおけるトピック間類似度を用いたトピック転換後の人気予測
3. 学会等名 第12回データ工学と情報マネジメントに関するフォーラム
4. 発表年 2020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
研究分担者	櫻井 保志 (Sakurai Yasushi) (30466411)	大阪大学・産業科学研究所・教授 (14401)	

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
---------	---------