

令和 6 年 6 月 25 日現在

機関番号：62615

研究種目：基盤研究(A)（一般）

研究期間：2019～2023

課題番号：19H01133

研究課題名（和文）構造抽出による自然言語ビッグデータへの高次高精度なデータマイニング技術の開発

研究課題名（英文）Efficient Text Big Data Mining Technology via Structure Extraction

研究代表者

宇野 毅明（Takeaki, Uno）

国立情報学研究所・情報学プリンシプル研究系・教授

研究者番号：00302977

交付決定額（研究期間全体）：（直接経費） 34,100,000円

研究成果の概要（和文）：形質から解析するSNS解析手法を多く構築した。1. マイクロクラスタリングとその時系列パターンを用いた、トピック間に単語と時系列情報の両方を考慮できる類似性を活用するクラスタリングの構築。2. LDAと文学的技術を融合した意味解釈手法の開発。巨大なツイート解析の意味づけを可能とした。3. 多様性を用いた文章集合の比較手法。炎上など多様性の低さに特徴を持つ事柄に関する類似性を効果的に測定できる。4. マイクロクラスタリングを用いたキーワード抽出手法。従来手法より局所的なトピックに的確に対応する。他にも多くのマイニングアルゴリズムやSNS分析結果で成果を出し、多くは査読付き論文として発表した。

研究成果の学術的意義や社会的意義

SNS分析を、その意味だけではなく、投稿時刻の分布や多様性など、人々の動きのパターンから行えるようになることで、意味の上からは観察しにくい、人々の群れとしての心理をひもとくような観察を行うことができるようになる。また、このような文章の集合、人々の群れの構造を元に、類似性やキーワード抽出を行うことで、今までとは質の異なる、直感により近づくと思えるような解を得られるようになった。また、マイニング技術と文学の読みの技術を融合する研究を行い、人文学の技術による深い理解を伴う解析技術の開拓をした。

研究成果の概要（英文）：We developed several social media analysis methods based on traits. 1. Construction of clustering that utilizes similarity considering both words and time-series information between topics, using micro-clustering and its time-series patterns. 2. Development of a semantic interpretation method that integrates LDA and techniques in literary, enabling the semantic analysis of massive tweet data. 3. A method for comparing sets of texts using diversity, effectively measuring the similarity of issues characterized by low diversity, such as online flaming. 4. A keyword extraction method using micro-clustering, which more accurately corresponds to local topics compared to conventional methods. In addition to these, we achieved results of many other mining algorithms and social media analysis, many of which have been published as peer-reviewed papers.

研究分野：アルゴリズム理論

キーワード：SNS解析 多様性 クラスタリング ソーシャルメディア データマイニング アルゴリズム 形質 類似性

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

近年の情報技術、自動推薦などによって、フィルターバブルや、病気や子育ての不安を解消するために絶え間ない検索を続けてしまうことなど、今までにはない新しい弊害が起きはじめている。これらの情報は、デマ情報やエセ科学もそうであるが、文章として、内容としては整合性があり、ある程度の支持も受けているために、自然言語解析などの意味理解、状況理解によるアプローチでは解決が難しい。この問題に対する有効な手立ては、個別の情報の文章理解ではなく、多くの人々の意見を“群れ”としての文章・意見の集合の流れや、全体像の把握をすることが効果的だろう。しかし、現在の自然言語解析の注目は1文の意味理解に集中している。

一方で、人間はそれほど深い意味をとらえずに、Web上の議論や意見のおおまかな傾向をとらえることができる。例えば、このトピックに関しては様々な意見がある、最近この事柄に関するツイートが急激に増加(バースト)している、この炎上が起こったとき、その周辺の事柄も注目されたり炎上したりした、といった具合である。

人間は深い意味解釈を伴わず、意見が投稿される頻度やその変化、いくつかのバーストしたトピックの関係性、類似した意見のまとまりなど、個々の意見や投稿などが作る意味的、時間的な局所構造を見て、それらの関係性を獲得していると考えられる。では、コンピュータにも、意味解釈を行わない議論の動きや趨勢の獲得ができるであろうか。我々は、様々なトピックのツイート数と、クラスタリングで得られるクラスタ数の関係を調べることで、デマ情報が抽出できることを確認した。クラスタやバーストに代表される意味的、時間的な局所構造を用いて、意見や議論の質と動きは獲得できるか、どこまで質を向上できるか、これが本課題で取り組む「問い」である。

### 2. 研究の目的

この課題の目的は、SNSや長大な文章などの大量のセンテンスが作る自然言語ビッグデータに対して、クラスタやバーストといった意味的、時間的な局所構造をモデル化、マイニングし、それに対して類型化や関連性などの高次なマイニングを行うことで、今までの解析技術では獲得できなかった、意見や議論の動きや質などといった、より高次の意味を獲得する解析手法の開発である。具体的には、

- (1) データ研磨[2]を発展させ、高精度な意味的局所構造を得ること
- (2) バーストやクラスタ等の局所構造に対する意味的に深い解析を行うこと
- (3) 局所構造の集合が持つ特性や関連性を調べ、それらと対応する高次の意味・概念(世論の動きなど)を獲得すること

を行う。データ研磨は、代表者が自身のCREST課題において開発した手法で、データの揺らぎを取り除き、内包する構造を明確化することにより、質の高いクラスタを抽出する方法である。このクラスタとバーストを用いて、SNSに展開される意見や記事を既存研究では決して得られない精度で分類し、多様性や質の変化を高い精度で獲得する手法を研究する。

また、本研究では、バーストの形質をモデル化する、賛成意見の多様性と反対意見の多様性を比べる、バーストと多様性の関係性を調べる、など既存研究では考慮されなかった方向からSNSや記事の解析手法を構築し、より高次の意味を獲得する技術を開発する。

### 3. 研究の方法

この課題は、自然言語ビッグデータに対する意味的、時間的な局所構造のモデル化と、上位のマイニングを用いた高次の意味を獲得する解析手法を開発する。到達目標は、大目標として、意味的、時間的な局所構造を軸とした、自然言語ビッグデータの高次な意味情報を獲得する手法体系の構築を目指している。本課題はその第一歩としての礎であり、モデルとアルゴリズムの開発を通じて、どのような局所構造と、どのような概念、議論の動き、炎上、感動や不安の増加や沈静化など、が対応するか、あるいは対応する見込みがあるのかを明らかにする。構築したアルゴリズムは実装を公開し、実際のSNSや記事データの解析が一般的に行えるようにする。

上記目標を達成するため、具体的に以下の4項目を研究する。

#### 1. 解析に資する意味的・時間的局所構造を精度良く見つける技術の開発

既存技術とデータ研磨を用いれば、バーストやクラスタは十分精度良く見つかる。しかし、言語データでは「Aである」「Aでない」のように「ない」の違いだけで意味が正反対になるなど、類似したものの中に対立構造があるなど、単なるクラスタリングでは不十分である。情報の拡散、変化のきっかけなど、クラスタやバーストからは見いだせない概念もある。この課題では、対立、賛成と言った概念的に大きな意味的に同質のものが作る意味的・時間的な局所構造をモデル化し、類似度や関連性を定義する。その元でクラスタリングやバースト検出を行うことで、精度の高い

発見を可能とする。

2. 多数の抽出した構造が作り出す、“群れ”としての関連性や高次の構造を見つけ出し、それらを可視化する手法の開発

クラスタ数とツイート数の比とデマ情報の関連のように、多数のクラスタやバーストの数や化、互いの関連性や密度などを調べ、多様性、自発性、拡散、沈黙化といった大きな概念との関連を網羅的に調べ上げる。これには、橋本(分担)が持つ、社会ネットワークの解析技術を適用する。

3. 開発した技術の実データで検証と、モデル開発に対するフィードバック

自然言語データは、他のセンサーデータなどとは異なり、人手による解析が機械よりも精度も質も非常に高い。そこで、開発手法の解を手手で検証することで、解が満たしていない要件や、見つけ損なっている解などの発見を行う。それにより、モデル化へとフィードバックし、より質の高いモデルを開発する。

4. 上記の技術に対する評価尺度、および評価方法の開発

モデルやアルゴリズムにはその評価が重要であるが、自然言語のような認知的な対象には、数理的に明確な評価が一般に難しい。本課題の開発技術は、意味解釈という認知プロセスを伴わないため、比較的数理との相性が良い。モデル化で用いた考え方は、そのまま解の尺度にも使える。例えば、SNSにおける現在の多様性を測る、変化度を測る、といったものに、クラスタの数や、スパイクの数を用いることができる。それら数値と、炎上や議論の潮目といった現象との関連を調べることで、これら現象としてしかとらえることのできなかった概念を、数値と関連づけて考えることができるようになる。アルゴリズムの評価も、教師ありベンチマークデータにおいて、それらの尺度でどれくらい良いかで評価することが可能になる。

4. 研究成果

本課題では、以上の研究計画に基づき、多くの成果を上げた。ここでは、その中から、基礎的なものをのぞき、大きな4つの成果を記する。

1つ目は、2-stage クラスタリングである。2-stage クラスタリングは、まずグラフクラスタリングを用いて Twitter (X) データからマイクロクラスターを抽出する第一段階と、それぞれのマイクロクラスターから時系列データを取得し、時系列クラスタリングを適用して coarse-grained なトピックを特定する第二段階の2つのステージから成るクラスタリング手法である。本手法は、意味的な類似性に加え、バーストの仕方のような時間軸の情報を用いることも可能となっており、従来手法よりも深い類似性に基づく解析が可能となっている。データ研磨技術を活用しており、大規模なデータに対して効率的に処理を行う能力に優れている。特に Tweets のような短いメッセージの集合のようなスパースなデータでも活用することが可能である。

2-stage クラスタリングの強みとして、データの特性を段階的に把握しながら分析を進めることができる点が挙げられる。第一段階で得られたマイクロクラスターに基づいて、より大きな文脈でのトピックの変遷を第二段階で明確にすることが可能となり、具体的なイベントや時事問題に対する人々の反応の変化を、時間の経過とともに追跡・分析することができる。本手法は、実際の大規模な Twitter データ (200 万から 600 万件の Tweets) に適用され、COVID-19 のワクチン接種や東京オリンピックなどの大きな社会イベントの前後でのツイートの変化や、それに伴う人々の認識を分析した。この方法は、速報ニュースへの反応と他のツイートやネット上のバズワードなどへの反応を効果的に区別する。計算効率の観点からも、きわめて高いパフォーマンスを示している。

2つ目は、ツイートのクラスタの深い意味解釈を行い、ツイートしている人々がどのような振る舞いをしているかを解析したものである。

2021 年 1~10 月に投稿された「ワクチン」を含む全ての日本語ツイート (800 万ユーザが投稿した 1.1 億ツイート) を収集した。まず、ツイートデータの前処理を行った。具体的には、絵文字の除去、形態素解析エンジン MeCab を用いた品詞への分割、ストップワード(「これ」「それ」「する」など)の除去、出現頻度が極端に低い単語の除去、を行った。ツイートの投稿頻度の時系列分析を行うと、職域接種が開始された 6 月から 8 月に多くのツイートが投稿されていた。この非定常性を取り除くため、各月から 10 万ツイートずつ、合計 100 万ツイートをランダム抽出した。次に、機械学習技術の 1 つであるトピックモデル (LDA: Blei et al. 2003) を用いることにより、100 万ツイートを 15 個のトピックに分類した。トピックモデルを使うと、単語の情報に基づいて膨大な数のツイートを自動的にトピックに分類できる。しかし、トピックモデルは分類されたトピックの意味 (どのような内容についてのトピックなのか) を解釈できない。そこで、各トピックからランダムに抽出した 1 万以上のツイート群を読みこみ、精査することによって 15 個のトピックの意味解釈を行い、「ワクチンの有効性」、「政治に関する意見」、「マスコミの報道に対する意見」などの 15 個のトピックに分類した。さらに、トピックの内容に基づいて人手でクラスタリングを行うことで、15 トピックを 4 つのテーマ (1. 個人的事柄、2. ニュース、3. 政治、4. 陰謀論とユーモア) のいずれかに振り分けた。結果、陰謀論に関するツイートは極めて少ないこと、職域接種、東京オリンピックの 2 つのイベントがツ

イートに与える影響が大きい(統計的に有意である)こと、などがわかった。

3つ目は、多様性を用いて、SNSなどの投稿の集合に対して類似性を定義し、従来とは異なる視点から、炎上のような多様性が低くなり議論がある種の健全性を失っている状態を検出し、あるいは人々の振る舞いの類似性を多様性の観点から明らかにしたものである。

具体的には、コメントに対する単語の多様性の解析からヤフーニュースのコメントの様相を分析した。ここでは単語の多様性とは意味的な多様性ではなく、単に文字として異なる単語がいくつ、どのような割合で入っているか、という意味である。コメントの単語の多様性は、コメント、意見、感想などの多様性を表していると考えられる。同じ意見でも、物事を説明している場合は、説明の仕方、根拠の付け方は多様であるため、多様性が高くなると考えられ、一方で感情的、紋切り型の意見は、感情を表す言葉は説明に比べると多様でないため、多様性が相対的に低くなると考えられる。2つのコメント集合があるとき、2つを1つに混ぜたら一般には多様性が上がると考えられるが、1つにしても多様性があがらなかつたら、両者は同じようなことに対して同じよう意見を述べている可能性が高いと考えられるだろう。このように、多様性、という普段はあまり見ない側面からコメント集合を見ることにより、キーワードなどからは得られない情報が獲得できると考えられる。具体的には、ベキ乗則当てはめ等の方法で多様性を図ることで、人が感じる直感に近い類似性を与える評価手法、かつ計算が極めて高速に行えるものを構築することができ、コロナ禍のニュースのコメントについて多くの知見を得ることができた。

4つ目は、マイクロクラスタリングを用いた、多様な事柄が話し合われている SNS 投稿の集合から、トピックとしてかたられているキーワードを的確に抽出する手法である。

まず、数多くある投稿をクラスタリングすることで、ある程度似ているものを集めて多数のマイクロクラスタを作り、各グループの中から重要な単語、つまり頻度が高い単語を見つける。ここでは各投稿の類似度として、それらの投稿に出現する単語集合がどの程度類似するかで判断する。クラスタリングにデータ研磨を用いることで、多数のマイクロクラスタを効率良く求める。マイクロクラスタはクラスタ内の投稿の出現単語が多く共通するようになるため、内容的にも似ている物が集まる確率が高い。そして、各クラスタで多く現れる、例えばクラスタ内の半数以上の投稿で現れる、単語をそのクラスタで語られているトピックに対応するキーワードであると考えられる。マイクロクラスタが類似性高い投稿を集めることから、通常のクラスタリングと異なり、このような方法でも効果的にキーワードが抽出できる。多くのクラスタでキーワードとなっている単語は、データ全体で重要なトピックに対応するキーワードになっていると考えられる。実際に、企業の炎上や、地震直後のツイートを分析した結果、通常の TF-IDF などに比べ、当時重要度が高かった単語を的確に抽出することに成功している。

## 5. 主な発表論文等

〔雑誌論文〕 計36件（うち査読付論文 29件 / うち国際共著 13件 / うちオープンアクセス 9件）

1. 著者名 Takako Hashimoto, David Lawrence Shepard, Tetsuji Kuboyama, Kilho Shin, Ryota Kobayashi, Takeaki Uno	4. 巻 77(5)
2. 論文標題 Analyzing temporal patterns of topic diversity using graph clustering	5. 発行年 2021年
3. 雑誌名 J. Supercomput	6. 最初と最後の頁 4375-4388
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 Kazuhiro Kurita, Kunihiro Wasa, Takeaki Uno, Hiroki Arimura	4. 巻 874
2. 論文標題 A constant amortized time enumeration algorithm for independent sets in graphs with bounded clique number	5. 発行年 2021年
3. 雑誌名 Theoretical Computer Science	6. 最初と最後の頁 32-41
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 Kazuhiro Kurita, Kunihiro Wasa, Hiroki Arimura, Takeaki Uno	4. 巻 303
2. 論文標題 Efficient enumeration of dominating sets for sparse graphs	5. 発行年 2021年
3. 雑誌名 Applied Mathematics	6. 最初と最後の頁 283-295
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 Alessio Conte, Mamadou Moustapha Kante, Takeaki Uno, Kunihiro Wasa	4. 巻 303
2. 論文標題 Maximal strongly connected cliques in directed graphs: Algorithms and bounds	5. 発行年 2021年
3. 雑誌名 Discrete Applied Mathematics	6. 最初と最後の頁 237-252
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 Murayama T, Wakamiya S, Aramaki E, and Kobayashi R	4. 巻 16(4)
2. 論文標題 Modeling the Spread of Fake News on Twitter	5. 発行年 2021年
3. 雑誌名 PLOS ONE	6. 最初と最後の頁 e0250419
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 Takako Hashimoto, Takeaki Uno, Yuka Takedomi, David Shepard, Toyoda Masashi, Naoki Yoshinaga, Masaru Kitsuregawa, Ryota Kobayashi	4. 巻 2021
2. 論文標題 Two-stage Clustering Method for Discovering People's Perceptions: A Case Study of the Covid-19 Vaccine from Twitter	5. 発行年 2021年
3. 雑誌名 Bigdata 2021	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 -

1. 著者名 Ryota Kobayashi, Patrick Gildersleve, Takeaki Uno, Renaud Lambiotte	4. 巻 2022
2. 論文標題 Modeling Collective Anticipation and Response on Wikipedia	5. 発行年 2022年
3. 雑誌名 Proceedings of the Fifteenth International AAAI Conference on Web and Social Media(ICWSM)	6. 最初と最後の頁 315-326
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 Huk, Maciej, Kilho Shin, Tetsuji Kuboyama, and Takako Hashimoto	4. 巻 2021
2. 論文標題 Random Number Generators in Training of Contextual Neural Networks.	5. 発行年 2021年
3. 雑誌名 In Intelligent Information and Database Systems: 13th Asian Conference, ACIIDS 2021	6. 最初と最後の頁 717-730
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 Hashimoto Takako, Shepard David Lawrence, Kuboyama Tetsuji, Shin Kilho, Kobayashi Ryota, Uno Takeaki	4. 巻 77
2. 論文標題 Analyzing temporal patterns of topic diversity using graph clustering	5. 発行年 2020年
3. 雑誌名 The Journal of Supercomputing	6. 最初と最後の頁 4375-4388
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s11227-020-03433-5	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 該当する

1. 著者名 Kusaba Akira, Hashimoto Takako, Shin Kilho, Shepard David Lawrence, Kuboyama Tetsuji	4. 巻 Volume 1
2. 論文標題 Unsupervised Clustering based on Feature-value / Instance Transposition Selection	5. 発行年 2020年
3. 雑誌名 Proceedings of 2020 IEEE REGION 10 CONFERENCE (TENCON)	6. 最初と最後の頁 1192-1197
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/TENCON50793.2020.9293922	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 該当する

1. 著者名 Shin Kilho, Okumoto Kenta, Shepard David, Kuboyama Tetsuji, Hashimoto Takako, Ohshima Hiroaki	4. 巻 Volume 2
2. 論文標題 A Fast Algorithm for Unsupervised Feature Value Selection	5. 発行年 2020年
3. 雑誌名 Proceedings of the 12th International Conference on Agents and Artificial Intelligence (ICAART2020)	6. 最初と最後の頁 203-213
掲載論文のDOI (デジタルオブジェクト識別子) 10.5220/0008981702030213	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 該当する

1. 著者名 Takako Hashimoto, Akira Kusaba, Dave Shepard, Tetsuji Kuboyama, Kilho Shin, Takeaki Uno	4. 巻 Volume 1
2. 論文標題 Twitter Topic Progress Visualization using Micro-clustering	5. 発行年 2020年
3. 雑誌名 Proceedings of the 9th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2020)	6. 最初と最後の頁 585-592
掲載論文のDOI (デジタルオブジェクト識別子) 10.5220/0009160805850592	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 該当する

1. 著者名 Adelani DI, Kobayashi R, Weber I, and Grabowicz PA.	4. 巻 25(9)
2. 論文標題 Estimating Community Feedback Effect on Topic Choice in Social Media with Predictive Modeling	5. 発行年 2020年
3. 雑誌名 EPJ Data Science	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) 10.1140/epjds/s13688-020-00243-w	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 該当する

1. 著者名 Kusaba Akira, Shin Kilho, Shepard Dave, Kuboyama Tetsuji	4. 巻 -
2. 論文標題 Predictive Nonlinear Modeling by Koopman Mode Decomposition	5. 発行年 2020年
3. 雑誌名 International Conference on Data Mining Workshops (ICDMW)	6. 最初と最後の頁 811-819
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/ICDMW51313.2020.00118	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 該当する

1. 著者名 Higuchi Naoya, Imamura Yasunobu, Mic Vladimir, Shinohara Takeshi, Hirata Kouichi, Kuboyama Tetsuji	4. 巻 LNCS 12440
2. 論文標題 Pivot Selection for Narrow Sketches by Optimization Algorithms	5. 発行年 2020年
3. 雑誌名 International Conference on Similarity Search and Applications (SISAP)	6. 最初と最後の頁 33-46
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-030-60936-8_3	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 該当する

1. 著者名 David Lawrence Shepard, Takako Hashimoto, Kilho Shin, Takeaki Uno, Tetsuji Kuboyama	4. 巻 -
2. 論文標題 Unsupervised Rumor Detection on Twitter using Topic Homogeneity	5. 発行年 2020年
3. 雑誌名 15th Annual International Conference of the Alliance of Digital Humanities Organizations(DH)	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 該当する



1. 著者名 Alessio Conte, Mamadou Moustapha Kante, Yota Otachi, Takeaki Uno, Kunihiro Wasa	4. 巻 818
2. 論文標題 Efficient enumeration of maximal k-degenerate induced subgraphs of a chordal graph	5. 発行年 2020年
3. 雑誌名 Theoretical Computer Science	6. 最初と最後の頁 2-11
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.tcs.2018.08.009	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 該当する

1. 著者名 Adrian Pino Angulo, Kilho Shin	4. 巻 49
2. 論文標題 Mrmr+ and Cfs+ feature selection algorithms for high-dimensional data.	5. 発行年 2019年
3. 雑誌名 Applied Intelligence	6. 最初と最後の頁 1954-1967
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s10489-018-1381-1	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 Shusuke Morimoto, Hajime Kawamukai, Kilho Shin	4. 巻 14
2. 論文標題 Prediction of Crime Occurrence using Information Propagation Model and Gaussian Process	5. 発行年 2019年
3. 雑誌名 14th AsiaJCIS 2019	6. 最初と最後の頁 80-87
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/AsiaJCIS.2019.000-2	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 Tianyu Zhang, Lei Zhu, Qian Zhao, Kilho Shin	4. 巻 -
2. 論文標題 Neural Networks Weights Quantization: Target None-retraining Ternary (TNT)	5. 発行年 2019年
3. 雑誌名 CoRR, December 2019	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 無
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 Kobayashi R, Kurita S, Kurth A, Kitano K, Mizuseki K, Diesmann M, Richmond BJ, Shinomoto S	4. 巻 10
2. 論文標題 Reconstructing Neuronal Circuitry from Parallel Spike Trains	5. 発行年 2019年
3. 雑誌名 Nature Communications	6. 最初と最後の頁 4468
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41467-019-12225-2	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 該当する

1. 著者名 Levakova M, Kostal L, Monsempes C, Lucas P, Kobayashi R	4. 巻 16
2. 論文標題 Adaptive integrate-and-fire model reproduces the dynamics of olfactory receptor neuron responses in moth	5. 発行年 2019年
3. 雑誌名 Journal of The Royal Society Interface	6. 最初と最後の頁 20190246
掲載論文のDOI (デジタルオブジェクト識別子) 10.1098/rsif.2019.0246	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 該当する

1. 著者名 Kostal L, Kobayashi R	4. 巻 100
2. 論文標題 Critical size of neural population for reliable information transmission	5. 発行年 2019年
3. 雑誌名 Physical Review E	6. 最初と最後の頁 050401(R)
掲載論文のDOI (デジタルオブジェクト識別子) 10.1103/PhysRevE.100.050401	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 該当する

1. 著者名 樋口直哉, 今村安伸, 久保山哲二, 平田耕一, 篠原武	4. 巻 13(1)
2. 論文標題 狭い16ビットのスケッチを用いた高速最近傍検索	5. 発行年 2020年
3. 雑誌名 情報処理学会論文誌数理モデル化と応用 (TOM)	6. 最初と最後の頁 13-22
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 無
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 Naoya Higuchi, Yasunobu Imamura, Takeshi Shinohara, Kouichi Hirata, Tetsuji Kuboyama	4. 巻 11996
2. 論文標題 Annealing by Increasing Resampling	5. 発行年 2019年
3. 雑誌名 ICPRAM (Revised Selected Papers)	6. 最初と最後の頁 71-92
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-030-40014-9_4	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 Fumiya Tokuhara, Shiho Okinaga, Tetsuhiro Miyahara, Yusuke Suzuki, Tetsuji Kuboyama, Tomoyuki Uchida	4. 巻 11
2. 論文標題 Using Label Information in a Genetic Programming Based Method for Acquiring Block Preserving Outerplanar Graph Patterns with Wildcards	5. 発行年 2019年
3. 雑誌名 2019 IEEE 11th International Workshop on Computational Intelligence and Applications (IWCIA)	6. 最初と最後の頁 95-100
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/IWCIA47330.2019.8955031	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 Naoya Higuchi, Yasunobu Imamura, Tetsuji Kuboyama, Kouichi Hirata, Takeshi Shinohara	4. 巻 LNCS 11919
2. 論文標題 Fast Filtering for Nearest Neighbor Search by Sketch Enumeration Without Using Matching	5. 発行年 2019年
3. 雑誌名 Australasian Conference on Artificial Intelligence	6. 最初と最後の頁 240-252
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-030-35288-2_20	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 Kilho Shin, Kenta Okumoto, David Lawrence Shepard, Tetsuji Kuboyama, Takako Hashimoto, Hiroaki Ohshima	4. 巻 12
2. 論文標題 A Fast Algorithm for Unsupervised Feature Value Selection	5. 発行年 2020年
3. 雑誌名 Proceedings of the 12th International Conference on Agents and Artificial Intelligence, ICAART (2)	6. 最初と最後の頁 203-213
掲載論文のDOI (デジタルオブジェクト識別子) 10.5220/0008981702030213	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 該当する

1. 著者名 Takako Hashimoto, Akira Kusaba, Dave Shepard, Tetsuji Kuboyama, Kilho Shin, Takeaki Uno	4. 巻 9
2. 論文標題 Twitter Topic Progress Visualization using Micro-clustering	5. 発行年 2020年
3. 雑誌名 Proceedings of the 9th International Conference on Pattern Recognition Applications and Methods (ICPRAM)	6. 最初と最後の頁 585-592
掲載論文のDOI (デジタルオブジェクト識別子) 10.5220/0009160805850592	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 Akira Kusaba, Tetsuji Kuboyama, Takako Hashimoto	4. 巻 10
2. 論文標題 Time Series Electricity Consumption Analysis using Non-negative Matrix Factorization	5. 発行年 2019年
3. 雑誌名 IEEE 10th International Conference on Awareness Science and Technology (iCAST)	6. 最初と最後の頁 1-6
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/ICAwST.2019.8923311	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 宇野毅明	4. 巻 177
2. 論文標題 データ研磨アルゴリズムの挙動と収束性の解析	5. 発行年 2020年
3. 雑誌名 第177回情報処理学会アルゴリズム研究会	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 無
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 宇野毅明	4. 巻 177
2. 論文標題 同型性を利用によるデータ研磨アルゴリズムの高速化手法	5. 発行年 2020年
3. 雑誌名 第177回情報処理学会アルゴリズム研究会	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 無
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 岩崎 幸子, 宇野 毅明	4. 巻 112
2. 論文標題 アピアランスを用いたクラスタリングにおける確信度の評価手法と実データによる検証	5. 発行年 2020年
3. 雑誌名 人工知能学会 第112回人工知能基本問題研究会	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 無
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 岩崎 幸子, 宇野 毅明	4. 巻 112
2. 論文標題 クラスタリング安定化手法を用いた芯クラスタの抽出と実データによる質の評価	5. 発行年 2020年
3. 雑誌名 人工知能学会 第112回人工知能基本問題研究会	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 無
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 Takako Hashimoto, Takeaki Uno, Tetsuji Kuboyama, Kilho Shin, Dave Shepard	4. 巻 -
2. 論文標題 Time Series Topic Transition Based on Micro-Clustering	5. 発行年 2019年
3. 雑誌名 Proceedings of IEEE International Conference on Big Data and Smart Computing, BigComp 2019	6. 最初と最後の頁 1-8
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 無
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 Alessio Conte, Takeaki Uno	4. 巻 51
2. 論文標題 New polynomial delay bounds for maximal subgraph enumeration by proximity search	5. 発行年 2019年
3. 雑誌名 Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019	6. 最初と最後の頁 1179-1190
掲載論文のDOI (デジタルオブジェクト識別子) 10.1145/3313276.3316402	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19件（うち招待講演 9件 / うち国際学会 10件）

1. 発表者名 東英寿, 久保山哲二
2. 発表標題 唐宋八大家古文の計量分析的考察—序、記、論の虚詞分析
3. 学会等名 『唐宋八大家研究』
4. 発表年 2022年

1. 発表者名 宇野 毅明
2. 発表標題 社会連携のためのファシリテーション講座 インタビューイング
3. 学会等名 九州大学公開講座（招待講演）
4. 発表年 2022年

1. 発表者名 宇野 毅明
2. 発表標題 社会連携のためのファシリテーション講座 相互理解とコミュニケーション
3. 学会等名 九州大学公開講座（招待講演）
4. 発表年 2022年

1. 発表者名 宇野 毅明
2. 発表標題 社会連携のためのファシリテーション講座 共通のテーマ設定
3. 学会等名 九州大学公開講座（招待講演）
4. 発表年 2022年

1. 発表者名 Hashimoto, T
2. 発表標題 "総合知に向けた人文社会科学データの観点から"
3. 学会等名 公開シンポジウム「総合知創出に向けた人文・社会科学のデジタル研究基盤構築の現在」
4. 発表年 2022年

1. 発表者名 徳永弘子, 久保山哲二, 木村敦, 武川直樹
2. 発表標題 共食における銘々膳と共同膳の形式が参加者の会話行動に及ぼす影響
3. 学会等名 電子情報通信学会技術研究報告121(363), pp.43-48, Jan. 2022
4. 発表年 2022年

1. 発表者名 Hashimoto, T.
2. 発表標題 Structuring Topics on Large-Scale Twitter for Discovering People's Perceptions
3. 学会等名 KJDB 2021
4. 発表年 2021年

1. 発表者名 Hashimoto, T.; Kusaba, A.; Shepard, D.; Kuboyama, T.; Shin, K. and Uno, T.
2. 発表標題 Twitter Topic Progress Visualization using Micro-clustering
3. 学会等名 the 9th International Conference on Pattern Recognition Applications and Methods the 9th International Conference on Pattern Recognition Applications and Methods (国際学会)
4. 発表年 2020年

1 . 発表者名 A. Kusaba, T. Hashimoto, K. Shin, D. L. Shepard and T. Kuboyama
2 . 発表標題 Unsupervised Clustering based on Feature-value / Instance Transposition Selection
3 . 学会等名 2020 IEEE REGION 10 CONFERENCE (TENCON) ( 国際学会 )
4 . 発表年 2020年

1 . 発表者名 David Lawrence Shepard, Takako Hashimoto, Kilho Shin, Takeaki Uno, and Tetsuji Kuboyama
2 . 発表標題 Unsupervised Rumor Detection on Twitter using Topic Homogeneity.
3 . 学会等名 The 15th Annual International Conference of the Alliance of Digital Humanities Organizations (DH2020) ( 国際学会 )
4 . 発表年 2020年

1 . 発表者名 Shin, K., Okumoto, K., Shepard, D. L., Kuboyama T., Hashimoto, T., and Ohshima
2 . 発表標題 A Fast Algorithm for Unsupervised Feature Value Selection
3 . 学会等名 The 12th International Conference on Agents and Artificial Intelligence (ICAART2020) ( 国際学会 )
4 . 発表年 2020年

1 . 発表者名 Sato, S., Hashimoto, T., and Shirota, Y.
2 . 発表標題 Evaluation for ESD (Education for Sustainable Development) to achieve SDGs at University
3 . 学会等名 The 2020 IEEE 10th International Conference on Awareness Science and Technology (iCAST) ( 国際学会 )
4 . 発表年 2020年



1. 発表者名 Hashimoto, T., Shin, K., Shepard, D. L., and Kuboyama T.
2. 発表標題 Indonesian Gender Equality Survey Analysis Using Feature Selection Based Clustering
3. 学会等名 The 2020 IEEE 10th International Conference on Awareness Science and Technology (iCAST) (国際学会)
4. 発表年 2020年

1. 発表者名 Hashimoto, T.
2. 発表標題 Efficient Text Big Data Mining Technology and Application
3. 学会等名 IEEE Computer Society India Symposium (招待講演) (国際学会)
4. 発表年 2020年

1. 発表者名 Hashimoto, T.
2. 発表標題 Efficient Text Big Data Mining Technology and Application
3. 学会等名 The 12th International Conference on Computational Collective Intelligence (ICCCI 2020) (招待講演) (国際学会)
4. 発表年 2020年

1. 発表者名 橋本 隆子
2. 発表標題 大規模SNSを対象としたテキストデータマイニング -集合行動解析とそのアプリケーション-
3. 学会等名 IEEE CS Japan/Tokyo Award記念講演会 (招待講演)
4. 発表年 2020年

1. 発表者名 橋本 隆子
2. 発表標題 BigData、Social Mining を通した新たなサービスモデルの可能性、特にコロナ後のダイバーシティ、グローバル化を睨んで
3. 学会等名 電子情報通信学会 研究会 KBSE・SC, 「ウィズコロナ/アフターコロナ時代の知能ソフトウェア工学とサービスコンピューティング」(招待講演)
4. 発表年 2020年

1. 発表者名 Hashimoto, T.
2. 発表標題 Detecting Collective Behavior on SNS Using Large-Scale Text Mining Techniques
3. 学会等名 The 17th INDICON 2020 (招待講演) (国際学会)
4. 発表年 2020年

1. 発表者名 Hashimoto, T.
2. 発表標題 Analyzing Collective Behavior on Social Networks Using Efficient Text Big Data Mining Technology.
3. 学会等名 IEEE PuneCon 2020 conference, (招待講演) (国際学会)
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	小林 亮太  (Kobayashi Ryota)  (70549237)	東京大学・大学院新領域創成科学研究科・准教授    (12601)	

6. 研究組織（つづき）

	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	久保山 哲二  (Kuboyama Tetsuji)  (80302660)	学習院大学・付置研究所・教授    (32606)	
研究分担者	橋本 隆子  (Hashimoto Takako)  (80551697)	千葉商科大学・商経学部・教授    (32504)	
研究分担者	申 吉浩  (Shin Yoshihiro)  (60523587)	学習院大学・付置研究所・教授    (32606)	

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			
イタリア	ピサ大学			
フランス	オーベル・ニュ・クレモンフェラン大学			