

平成 30 年 5 月 19 日現在

機関番号：34304

研究種目：基盤研究(C) (一般)

研究期間：2014～2017

課題番号：26330351

研究課題名(和文)大規模ソーシャルメディア分析に基づくトレンド早期発見技術に関する研究

研究課題名(英文)Early Detection of Buzzwords Based on Large-scale Social Media Analysis

研究代表者

中島 伸介 (NAKAJIMA, Shinsuke)

京都産業大学・コンピュータ理工学部・教授

研究者番号：90399535

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)：本研究課題は、ソーシャルメディアを効率的かつリアルタイムに近い状態で解析することで、世の中に広まる前のトレンド(流行語)をいち早く発見することを目指したものである。解析対象ソーシャルメディアとしては、1,113万ブログサイトにて投稿された1億4,178万件のブログ記事を利用した。流行語の早期発見手法としては、コミュニティ間の話題の拡散分析に基づく方法およびブロガー先読み度分析技術に基づく方法について取り組んだ。また、世の中のトレンドおよび流行語、さらには各個人の興味トレンド分析し、これらをWeb広告、料理レシピ、コスメ、クチコミなど各種情報推薦に活用する手法についても検討を行った。

研究成果の概要(英文)：In this project, we tried to develop a method for early detection of “gradual buzzwords” by analyzing time-series data of blog entries. We utilized 141 million blog entries from 11 million blog websites. Then, we studied two methods for detecting gradual buzzwords. One is a method which evaluates bloggers' buzzword prediction ability by analyzing how early bloggers mentioned past buzzwords. Another is a method based on observing the process in which certain topics grow to become major buzzwords and determining the key indicators that are necessary for their early detection. Moreover, we studied how to utilize trends and buzzwords of the world and also trends of individual interests in order to improve the accuracy for recommender systems of Web advertisement, cooking recipe, cosmetics and the reviews.

研究分野：情報工学

キーワード：Webマイニング トrend分析 流行語早期発見

### 1. 研究開始当初の背景

インターネットの発達や、タブレット端末およびスマートフォンなどの普及と共に、多くの人々が情報発信を行うようになった。平成23年度情報通信白書(総務省)によれば、平成23年時に行った調査において42.9%の回答者がソーシャルメディアを利用していると答えている。この数値は10代に限れば71.7%となっており、世代の交代が進むにつれて今後さらに利用率が増えることが予想される。このような一般ユーザが発信する情報には、マスメディアが捉えきれないような世の中の動きに関する情報(ロングテール)が数多く隠されていると考えられる。実際、企業の中には、自社製品やサービスに関するクチコミ情報の分析を行う企業も数多く存在し、商品の改良やサービスの改善、さらには新規企画の参考情報として利用している。したがって、大規模ソーシャルメディアコンテンツを、効果的に分析することで、世の中の動向を精度良く捉えるための技術が求められているといえる。

### 2. 研究の目的

近年、爆発的に広まっているソーシャルメディアには、マスメディアが捉えきれないような世の中の動き(ロングテール)が数多く隠されているという仮説のもと、これらを効率的かつリアルタイムに近い状態で解析することで、世の中に広まる前のトレンド(流行語)をいち早く発見することを目指す。さらに、世の中のトレンドおよび流行語、および各個人の興味トレンド分析し、これらをWeb広告、料理レシピ、コスメ、クチコミなど各種情報推薦に活用する手法についても検討を行う。流行語の早期発見手法としては、コミュニティ間の話題の拡散分析に基づく方法およびプロガー先読み度分析技術に基づく方法について取り組む。解析対象ソーシャルメディアとしては、1,113万ブログサイトにて投稿された1億4,178万件のブログ記事(twitterを含む)を対象とする。最終的にはトレンド早期発見技術を開発し、大規模実データをを用いた実験的評価を行う。

### 3. 研究の方法

(1) コミュニティ間の話題の拡散分析に基づく流行語早期発見技術の開発では、ある話題がコミュニティ間で拡散する状況进行分析することにより、流行語を早期発見するための技術開発を目指す。例えば、「Facebook」という言葉は、当初「インターネット」や「SNS」等のように「Facebook」と関連の深いコミュニティにて独占的に語られていたものが、その他一般的なコミュニティにおいても広まり語られるようになったと考えられる。したがって、対象となる話題語と関連の深いコミュニティの自動判定手法、およびコミュニティ間の話題の拡散状況を精度良く検出する手法、について明らかにすると共

に、本方式による流行語判定技術の開発および実験的評価を行う。

(2) プロガー先読み度分析に基づく流行語早期発見技術の開発では、過去の投稿履歴から先読み能力の高いプロガーを分析することで流行語を早期発見するための技術開発を目指す。例えば、「新しいiPhoneの機能」について、いち早く言及することが多い情報発信者は、今後発表されるiPhoneの機能についてもいち早く言及することが期待できる。したがって、既出のメジャーな流行語の分析に基づく流行語成長過程分析手法、および流行語成長過程および対象の話題と各記事の関連度分析に基づくプロガーの先読み度分析手法、について明らかにすると共に、本方式によるプロガー先読み度分析技術の開発および実験的評価を行う。

(3)トレンド早期発見技術の開発および各種情報推薦への応用では、コミュニティ間の話題の拡散分析技術(直接的流行語発見手法)およびプロガー先読み度分析技術(間接的流行語発見手法)を融合することによるトレンド早期発見技術を開発すると共に、Web広告、料理レシピ、コスプレビューなど各種情報推薦への応用について検討する。

### 4. 研究成果

(1) 本研究では流行語に対していち早く反応する傾向を有する先読みプロガーの判定を目的としている。そこで流行語の流行時の話題が投稿された時期(成長期間)の算出についてより詳細に述べ、提案手法によって、対象流行語の流行時の内容を早期かつ、的確に捉えることができる先読みプロガーを発見が可能かについて実験を行った。提案手法の手順としては、対象プロガーの抽出とカテゴリ分類、学習データとなるシード語(過去の流行語)の抽出とカテゴリ分類、対象シード語(過去の流行語)の分析、および対象プロガーの先読み能力の評価という流れとなる。シード語(過去の流行語)の抽出では、ある程度重要なキーワードに絞る必要があるため、シード語候補毎にブログ投稿数を調べ、該当シード語候補の過去の投稿数に対し、必要に応じて移動平均を算出し、投稿数のピークを確認する。このピークを迎えた時点が社会的認知の最も高くなった時点であるといえる。このピーク以降の期間Tにおける投稿数の累計が非常に少なくなっている場合には、このシード語候補はピーク後に世間から忘れ去られるようなキーワードであると考えられるため、ピーク以降も投稿数があまり減少しないようなシード語候補を社会的な影響度が高いキーワードであると判断し、シード語として認定する(図1参照)。

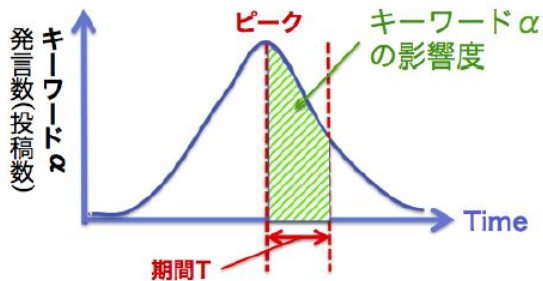


図1 影響度の概念図

各シード語に対するブロガー先読み度の算出方式については、シード語の投稿数のピーク時の話題に対して、類似度の高い記事を成長期間内でより早くから言及しているブロガーの先読み度を高く算出するようにした。(図2参照)

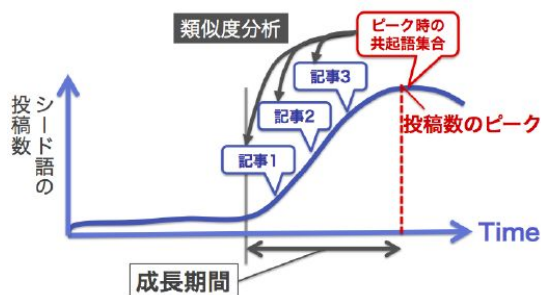


図2 ブロガー先読み度判定のためのブログ記事の類似度分析

以上のような手順により、ブロガー先読み度算出について実験を行い、ブロガー先読み度の計算対象となる記事の選別において、対象シード語について言及された記事がどうかを考慮に入れる事で、適切に先読みブロガーが発見されることを確認した。

(2)本研究では、先読みブロガーから流行語候補を抽出する手法を提案した。具体的には「先読み度の高い先読みブロガーが投稿したキーワード程、流行語候補である事が期待できる」という仮定に基づき、先読み度の高い先読みブロガーが投稿するブログ記事より、先読み度の高さを考慮して流行語候補を抽出する事を目指した。なお、流行は時間の流れとともに変化していくため、現時点において、流行語候補となりうるのか、過去の時点まで遡って、時系列の変化と共に、先読みブロガーが投稿する頻度が増加しているのか、また増加がその過去の流行語と比較して、十分な増加傾向にあると言えるのかを調べる事で流行語候補の判定を行った。例えば、ブックカテゴリの先読みブロガー A が投稿した記事内において、小説 X について書いていたとしても、小説 X が流行するとは限らない。しかし、ブックカテゴリの先読みブロガー群において、先読みブロガー A

だけでなく、他の先読みブロガーも小説 X について投稿しているならば、小説 X は、流行する可能性があると考えられる。ただし、単純に、短い期間において、ブックカテゴリの先読みブロガー群が、小説 X を多数投稿しているだけでは、小説 X を流行語候補とみなす事はできない。流行語候補と判定するためには、ある一定期間毎に、ブックカテゴリの先読みブロガー群の、小説 X に対する投稿傾向を分析したとき、過去の流行語と同じような、増加傾向であると判定する必要がある。以上より、先読み度の高さを考慮してキーワードを抽出し、対象カテゴリ内の先読みブロガー群において、ある一定期間毎の対象キーワードの投稿傾向を調べ、過去の流行語の増加傾向と比較することにより、対象キーワードを流行語候補と判定し、抽出する手法を提案した。具体的には、流行語毎に先読み度で上位ブロガーリストを作成し、リストに出現した回数を用いる手法(a)と、先読み度の合計を用いる手法(b)、記事数の影響を考慮した手法(c)、出現したリストのカバレッジを考慮した手法(d)とそれらの要素を統合した手法(e)を提案した。

映画、TV番組、スマホといった3つのカテゴリに対して、先読みブロガーの抽出精度を比較検証した。ブログ記事数とブロガーの熟知度といった2つのベースライン手法と比べて、3つのカテゴリとも精度の向上を達成できおり、提案手法の有効性を示せた(図3)。5つの提案手法に関しては、提案手法(a)と提案手法(e)が同程度の高い精度を達成した(図4)が、提案手法(a)で算出したブロガーのスコアは整数であるのに対して、提案手法(e)での算出値は実数であるため、提案手法(e)を用いると、同順位のブロガーを減らすことができ、ブロガーをより区別できると思われる。

カテゴリ	ブロガー数	ブログ記事数	熟知度	提案手法(a)
映画	8	12.5% (1/8)	0% (0/8)	62.5% (5/8)
TV番組	7	14.3% (1/7)	28.6% (2/7)	28.6% (2/7)
スマホ	6	50.0% (3/6)	16.7% (1/6)	66.7% (4/6)
平均	7	25.6%	15.1%	52.6%

図3 先読みブロガーの抽出精度(1)

カテゴリ	提案手法(a)	提案手法(b)	提案手法(c)	提案手法(d)	提案手法(e)
映画	62.5% (5/8)	50.0% (4/8)	37.5% (5/8)	62.5% (5/8)	62.5% (5/8)
TV番組	28.6% (2/7)	28.6% (2/7)	42.9% (3/7)	14.3% (1/7)	28.6% (2/7)
スマホ	66.7% (4/6)	50.0% (3/6)	50.0% (3/6)	66.7% (4/6)	66.7% (4/6)
平均	52.6%	42.9%	43.5%	47.8%	52.6%

図4 先読みブロガーの抽出精度(2)

(3)トレンド早期発見技術の各種情報推薦への応用として、Web 広告、料理レシピ、コスメレビューなど各種情報推薦への応用について検討した。Web 広告推薦技術への応用としては、ユーザ個人の Web 閲覧履歴の分析から個人の興味トレンドを精度良く分析し、その時々々の興味に合致する広告推薦の実現を目指している。この際、長期的閲覧履歴と短期的閲覧履歴を分析することで、ユーザの長期トレンドおよび短期トレンドを区別して抽出する手法を提案している。この手法により単一区間を分析する手法に比べて予測精度が高められることを確認した。料理レシピ推薦への応用としては、ユーザの気分(ムード)に適切なレシピを推薦しようとするシステムにおいて、ユーザの気分のトレンド分析への応用を検討した。コスメレビュー分析によるレビューおよびアイテム推薦への応用としては、レビューが書いたレビュー履歴を分析することでレビューの興味トレンド分析を行い、これをレビュー間の類似度判定に応用する手法について検討した。これに関連し、レビュー閲覧者の閲覧履歴データを取得することで、閲覧者の興味トレンドを抽出する手法についても検討している。その他、本研究のトレンド早期発見技術は、ホテルレビュー推薦や、Serendipity 情報推薦への応用についても活用可能であると考えており、本研究期間終了後においても他の研究テーマにおいても広く活用していく予定である。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

##### [雑誌論文](計5件)

Yuuki Matsunami, Mayumi Ueda, Shinsuke Nakajima, Takeru Hashikami, John O' Donovan, and Byungkyu Kang, "Mining Attribute-Specific Ratings from Reviews of Cosmetic Products", Transactions on Engineering Technologies (International MultiConference of Engineers and Computer Scientists 2016), pp.101-114, Springer, 2017. (査読有)

Mayumi Ueda, Natsuhiko Takata, Yukitoshi Morishita, and Shinsuke Nakajima, "Utilizing Various User Moods for Automatic Recipe-Metadata Generation", Transactions on Engineering Technologies (International MultiConference of Engineers and Computer Scientists 2016), pp.115-129, Springer, 2017. (査読有)

Jianwei Zhang, Seiya Tomonaga, Shinsuke Nakajima, Yoichi Inagaki, and Reyn Nakamoto, "Prophetic Blogger

Identification Based on Buzzword Prediction Ability", International Journal of Web Information Systems, Vol. 12, No. 3, pp.267-291, August 2016.

(査読有)

Seiya Tomonaga, Shinsuke Nakajima, Yoichi Inagaki, Reyn Nakamoto, Jianwei Zhang : "Analyzing Early Mentioning of Past Buzzwords for Determination of Bloggers' Buzzword Prediction Ability", Transactions on Engineering Technologies, International MultiConference of Engineers and Computer Scientists 2014, pp.353-367, Springer Netherlands, Jan. 2015. (査読有)

Mayumi Ueda, Shinsuke Nakajima : "Cooking Recipe Recommendation Method Focusing on the Relationship Between User Preference and Ingredient Quantity", Transactions on Engineering Technologies, International MultiConference of Engineers and Computer Scientists 2014, pp.385-395, Springer Netherlands, Jan. 2015. (査読有)

##### [学会発表](計27件)

Jianwei Zhang, Yoichi Inagaki, Reyn Nakamoto and Shinsuke Nakajima, "Estimating Bloggers' Prediction Ability on Buzzwords and Categories", Proc. IAENG International MultiConference of Engineers and Computer Scientists (IMECS 2018), pp.393-398, Hong Kong, China, March 2018. 松波友稀, 上田真由美, 中島伸介, 各コスメアイテム分類に対する評価表現辞書構築方法に関する考察, 第10回データ工学と情報マネジメントに関するフォーラム(DEIM Forum 2018) D1-3, 2018年3月。

山口由莉子, Siriaraya Panote, 森下民平, 稲垣陽一, 中本レン, 張建偉, 青井順一, 中島伸介, Web 広告推薦のための長期的・短期的興味を考慮したユーザの潜在的興味分析方式, 第10回データ工学と情報マネジメントに関するフォーラム(DEIM Forum 2018) C5-1, 2018年3月。

<DEIM2018 学生プレゼンテーション賞> 澤田礼我, Siriaraya Panote, 堀篤史, 山口由莉子, 森下民平, 稲垣陽一, 中本レン, 張建偉, 中島伸介, 潜在的興味分析に基づく詳細な興味グループ判定方式の提案, 第10回データ工学と情報マネジメントに関するフォーラム(DEIM Forum 2018) P8-2, 2018年3月。

Panote Siriaraya, Yuriko Yamaguchi, Mimpai Morishita, Yoichi Inagaki, Reyn Nakamoto, Jianwei Zhang, Junichi Aoi,

Shinsuke Nakajima, "Using categorized web browsing history to estimate the user's latent interests for web advertisement recommendation", The First IEEE Workshop on Human-Machine Collaboration in BigData (HMData2017) co-located with IEEE Bigdata 2017, pp.4347-4352, December 2017.

Yuki Matsunami, Mayumi Ueda and Shinsuke Nakajima, "Tag Recommendation Method for a Cosmetic Review Recommender System", The 19th International Conference on Information Integration and Web-based Applications & Services(iiWAS2017), pp.166-170, December 2017.

松波友稀, 上田真由美, 中島伸介, コスメアイテムに対する項目別スコアを用いた効果タグ推薦手法の提案. 第10回Webとデータベースに関するフォーラム(WebDB Forum 2017), DE2017-22(2017-09), pp.139-144, 2017. <WebDB2017 学生奨励賞>

Junya Yamazaki, and Shinsuke Nakajima, Serendipity-Oriented Recommender System Considering Product Awareness in Communities, International MultiConference of Engineers and Computer Scientists 2017, ICICWS2017 pp461-465, March 2017.

Yuki Matsunami, Asami Okuda, Mayumi Ueda, and Shinsuke Nakajima, User Similarity Calculating Method for Cosmetic Review Recommender System, International MultiConference of Engineers and Computer Scientists 2017, ICDMA2017 pp312-316, March 2017.

山口由莉子, 森下民平, 稲垣陽一, 中本レン, 張 健偉, 青井順一, 中島伸介, Web 広告推薦のための閲覧カテゴリ情報を用いたユーザの潜在的興味分析方式, 第9回データ工学と情報マネジメントに関するフォーラム(DEIM Forum 2017) B2-3, 2017年3月

松波友稀, 上田真由美, 中島伸介, コスメアイテムに対する評価項目別レビュー自動スコアリング方式の開発, 第9回データ工学と情報マネジメントに関するフォーラム(DEIM Forum 2017) B5-3, 2017年3月

堀 篤史, 山口由莉子, 森下民平, 稲垣陽一, 中本レン, 張 健偉, 中島伸介, Web空間におけるユーザ行動履歴および感情を考慮した潜在的興味分析の検討, 第9回データ工学と情報マネジメントに関するフォーラム(DEIM Forum 2017) P6-3, 2017年3月

奥田麻美, 松波友稀, 上田真由美, 中島伸介, コスメリビュー共有システムのための類似ユーザ判定手法, 第9回データ

工学と情報マネジメントに関するフォーラム(DEIM Forum 2017) P8-5, 2017年3月

梅崎準菜, 松波友稀, 上田真由美, 中島伸介, ユーザの気分と状況を考慮したネイルレシピ検索方式の提案, 第9回データ工学と情報マネジメントに関するフォーラム(DEIM Forum 2017) P8-6, 2017年3月

Yuriko Yamaguchi, Mimpei Morishita, Yoichi Inagaki, Reyn Nakamoto, Jianwei Zhang, Junichi Aoi, and Shinsuke Nakajima, "Web Advertising Recommender System Based on Estimating Users' Latent Interests", The 18th International Conference on Information Integration and Web-based Applications & Services(iiWAS2016), pp.44-51, 2016.11.

Mayumi Ueda, Yukitoshi Morishita, Tomiyo Nakamura, Natsuhiko Takata, and Shinsuke Nakajima, "Recipe Recommender System by Considering User's Moods", The 18th International Conference on Information Integration and Web-based Applications & Services(iiWAS2016), pp.474-478, 2016.11.

John O'Donovan, Shinsuke Nakajima, Tobias Höllerer, Mayumi Ueda, Yuuki Matsunami, Byungkyu Kang, A Cross-Cultural Analysis of Explanations for Product Reviews, Proceedings of the Joint Workshop on Interfaces and Human Decision Making for Recommender Systems (IntRS 2016) co-located with ACM Conference on Recommender Systems (RecSys 2016), pp55-58, September 2016.

Yuuki Matsunami, Mayumi Ueda, Shinsuke Nakajima, Takeru Hashikami, Sunao Iwasaki, John O'Donovan, and Byungkyu Kang, Explaining Item Ratings in Cosmetic Product Reviews, International MultiConference of Engineers and Computer Scientists 2016, ICICWS2016 pp392-397, March 2016.

Natsuhiko Takata, Mayumi Ueda, Yukitoshi Morishita, and Shinsuke Nakajima, Automatic Recipe Metadata Generating Method by Considering Users' Various Moods, International MultiConference of Engineers and Computer Scientists 2016, ICICWS2016 pp413-418, March 2016.

松波友稀, 上田真由美, 中島伸介, 階上 猛, 岩崎 素直, John O'Donovan, Byungkyu Kang, コスメアイテム評価表現辞書を用いた評価項目別レビュー自動スコアリング方式, 第8回データ工学と

- 情報マネジメントに関するフォーラム (DEIM Forum 2016) B1-1, 2016年3月
- 21 山口 由莉子, 森下 民平, 稲垣 陽一, 中本 レン, 張 建偉, 青井 順一, 中島 伸介, ユーザの潜在的興味に基づく Web 広告推薦方式の検討, 第 8 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2016) B1-2, 2016年3月
- 22 高田 夏彦, 上田 真由美, 森下 幸俊, 中島 伸介, 各種気分データに基づき自動付与された未知レシメタデータの妥当性評価, 第 8 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2016) E2-2, 2016年3月
- 23 Jianwei Zhang, Seiya Tomonaga, Shinsuke Nakajima, Yoichi Inagaki, and Reyn Nakamoto, "Finding Prophets in the Blogosphere: Bloggers Who Predicted Buzzwords Before They Become Popular", Proc. The 17th International Conference on Information Integration and Web-based Applications & Services (iiWAS 2015), pp. 100-109, Brussels, Belgium, December 2015.
- 24 松波 友稀, 上田 真由美, 中島 伸介, コスメアイテムの使用感および嗜好度判定を目的としたレビュー分析手法の提案, 第 7 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2015) P3-1, 2015年3月. <DEIM2015 優秀インタラクティブ賞>
- 25 山口 由莉子, 朝永 聖也, 中島 伸介, 稲垣 陽一, 中本 レン, 森下 民平, パーチェスファネルに基づく Web 広告推薦手法の提案, 第 7 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2015) P1-4, 2015年3月.
- 26 田川 和幸, 朝永 聖也, 中島 伸介, 稲垣 陽一, 中本 レン, 張 建偉 先読みブロガーによる投稿記事の分析に基づく流行語予測手法の提案, 第 7 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2015) P3-2, 2015年3月.
- 27 朝永聖也, 中島伸介, 稲垣陽一, 中本レン, 張 建偉. 過去の流行語を学習データに用いた流行先読みブロガー発見手法の提案, Web とデータベースに関するフォーラム (WebDB Forum 2014), 情報処理学会シンポジウムシリーズ, A-2-2, 2014年11月. (査読あり) <学生奨励賞>

張 建偉 (ZHANG, Jianwei)  
岩手大学・理工学部・准教授  
研究者番号: 20635924

## 6. 研究組織

### (1) 研究代表者

中島 伸介 (NAKAJIMA, Shinsuke)  
京都産業大学・コンピュータ理工学部・教授  
研究者番号: 90399535

### (2) 研究分担者