

令和元年6月26日現在

機関番号：32689

研究種目：基盤研究(C)（一般）

研究期間：2016～2018

課題番号：16K00423

研究課題名（和文）ソーシャルメディアからの構造的知識の抽出と投稿意図分析

研究課題名（英文）Extraction of structural knowledge and analysis of posting intention from social media

研究代表者

岩井原 瑞穂（IWAHARA, MIZUHO）

早稲田大学・理工学術院（情報生産システム研究科・センター）・教授

研究者番号：40253538

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）：時間経過を考慮した構造的知識の抽出では、編集履歴を有する記事集合から、バースト的に編集が行われた文やフレーズを抽出する手法を開発した。数百からなる比較的大きな記事集合に対し、バースト的編集に関連を持つ語句を抽出する方法を開発し、記事集合の発展とともに重要語句が変遷することが確認できた。

ソーシャルメディアにおける投稿の意図分析では、投稿の文章に対し、センチメント分析を適用して、正負の極性ならびに主観的か客観的か、および文の長さ等の多次元データを求めた。これからユーザごとのセンチメントの分布が7～8程度の潜在的な投稿タイプに分類でき、ツイートの筆者特定精度を向上に適用できることが分かった。

研究成果の学術的意義や社会的意義

本研究は、ソーシャルメディアからの情報抽出において、新たにWikipediaなどの編集履歴からのキーフレーズ抽出手法を開発した。編集作業は人間の著者による事項の取捨選択が行われ、サーチエンジンにおける検索語とは異なる、より要約された事項の発展度合いを検出できることが分かった。

またツイートなどの短い書き込みを扱う多様なソーシャルメディアが展開されているが、短い書き込みではより鮮明な感情の表出が検出できる。書き込み集合における正負の感情の分布が、ユーザごとに違いが生じることが分かったが、これを筆者特定やレビューの補正などに用いることができる。

研究成果の概要（英文）： We developed a method to extract key sentences and phrases that are correlated with bursting edits from versioned articles. Key phrases related to bursts happening in edit histories of hundreds of articles can be detected, which can be used to find how significant phrases are evolving over time.

For analysis of posting intention in social media, we conducted collection of SNS posts and carried out sentiment analysis, and obtained multi-dimensional feature vectors that summarize users' sentiment orientations. Seven to eight latent posting styles are appeared to be reasonable to explain users' posting styles, which can be applied to authorship attribution on tweets.

研究分野：情報工学

キーワード：ソーシャルメディア ユーザ行動分析 テキストマイニング 意味的関連性 時系列分析 クラスターング

## 様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

### 1. 研究開始当初の背景

#### (1) 時間経過を考慮した構造的知識の抽出

代表的な知識蓄積型ソーシャルメディアである Wikipedia は、ユーザの投稿に基づく巨大な百科事典であるとともに、リンク関係やカテゴリ、infobox など計算機利用が容易な構造的データを含み、知識ベースとして抽出され検索サービスに活用されるなど、データグラフとしての応用が進んでいる。一方、Wikipedia には、各記事の過去のバージョンが編集履歴として蓄積・公開されているため、時系列データとしても利用できる。本テーマでは、文脈および時間の両面を考慮した新たな情報抽出技術の開発と、その応用展開を目的とする。

#### (2) ソーシャルメディアにおける投稿の意図分析

ソーシャルメディアは多数のユーザの投稿で成立しているが、ユーザがどのような目的や動機で投稿を行っているかの推定は、コンテンツ推薦やユーザ推薦、広告、サービスの改良等の幅広い応用がある。しかし、ユーザが自身のプロフィール情報や投稿を公開することは、プライバシー上のリスクがある。近年は不用意な情報の公開にユーザは慎重になってはいるが、ある程度の情報公開を行わないと相手に認知され共感してもらう機会が減るといった不利益がある。ユーザはリスクと利益のバランスを考慮して投稿行動を行っていると考えられるが、ソーシャルメディアの形式や利用形態は日々変化しており、ユーザの投稿行動も変化を続けていると考えられ、投稿の動機分析を継続的に行うことにより、ソーシャルメディアに対するユーザの態度の変化を定量的に特徴づけられると考えられる。

### 2. 研究の目的

Wikipedia や Facebook に代表されるソーシャルメディアは、多数のユーザの投稿で成立するユーザ参加型メディアである。本課題では、ソーシャルメディアの先導的応用の観点から次の2つのテーマについて研究を行う。

- (1) 時間経過を考慮した構造的知識の抽出では、編集履歴を有する記事集合から文単位の派生関係を求めることにより、文脈的あるいは時間的に重要なフレーズを抽出する技術を開発し、実体リンクや Wikipedia リストに含まれるべき記事の判定など、構造的知識への抽出への応用を行う。
- (2) ソーシャルメディアにおける投稿意図分析では、ソーシャルメディアにおいてユーザが投稿を一般に公開する場合に、どのような意図に基づいているかをモデル化し、外部から観測可能な情報のみにより投稿意図を推定する統計モデルの構築を行う。

### 3. 研究の方法

(1) 時間経過を考慮した構造的知識の抽出では、編集履歴を有する記事集合から文や節などの単位で意味的な関係性を評価することにより、文脈的あるいは時間的に重要なフレーズを抽出する。時間的側面では、週や月といった単位でバースト的な編集が行われているフレーズを抽出する方法として、編集履歴からフレーズを選択する評価関数を定義する。これを記事集合の履歴に適用してバーストの検出を行う。また、実体リンクについては、Wikipedia のカテゴリ階層も反映した意味的関連性を定義して、リンクの精度を上げることを試みる。さらに記事間のリンク先として、記事全体ではなく、より関連の深い節を求める応用にも適用する。

(2) ソーシャルメディアにおける投稿の意図分析では、ソーシャルメディアにおいてユーザが投稿を一般に公開する場合に、どのような意図に基づいているかをモデル化し、外部から観測可能な情報のみにより投稿意図を推定する統計モデルの構築を行う。具体的には、まずユーザプロフィールと公開投稿の収集を行ない、投稿の意図を手動で分類し、階層的に集約することにより、公開投稿の意図分類モデルを構築する。投稿頻度やプロフィールで5つの意図分類のうちどれと相関性があるかを統計分析により求める。また、投稿の文章に対し、センチメント分析を適用して、正負の極性ならびに主観的か客観的かの2軸で分類を行なう。ユーザ個人ごとのセンチメントの分布に対し、クラスタリングを行い典型的な類型に分類する。

### 4. 研究成果

時間経過を考慮した構造的知識の抽出では、編集履歴を有する記事集合から、バースト的に編集が行われた文やフレーズを抽出する手法を開発した。数百からなる比較的大きな記事集合に対し、バースト的編集に関連を持つ語句を抽出する方法を開発し、記事集合の発展とともに重要語句が変遷することが確認できた。

ソーシャルメディアにおける投稿の意図分析では、投稿の文章に対し、センチメント分析を適用して、正負の極性ならびに主観的か客観的か、および文の長さ等の多次元データを求めた。これからユーザごとのセンチメントの分布が7-8程度の潜在的な投稿タイプに分類でき、ツイートの筆者特定の精度を向上に应用できることが分かった。

We developed a method to extract key sentences and phrases that are correlated with bursting edits from versioned articles. Key phrases related to bursts happening in

edit histories of hundreds of articles can be detected, which can be used to find how significant phrases are evolving over time.

For analysis of posting intension in social media, we conducted collection of SNS posts and carried out sentiment analysis, and obtained multi-dimensional feature vectors that summarize users' sentiment orientations. Seven to eight latent posting styles are appeared to be reasonable to explain users' posting styles, which can be applied to authorship attribution on tweets.

#### 研究成果の学術的意義や社会的意義

本研究は、ソーシャルメディアからの情報抽出において、新たに Wikipedia などの編集履歴からのキーフレーズ抽出手法を開発した。編集作業は人間の著者による事項の取捨選択が行われ、サーチエンジンにおける検索語とは異なる、より要約された事項の発展度合いを検出できることが分かった。

またツイートなどの短い書き込みを扱う多様なソーシャルメディアが展開されているが、短い書き込みではより鮮明な感情の表出が検出できる。書き込み集合における正負の感情の分布が、ユーザごとに違いが生じることが分かったが、これを筆者特定やレビューの補正などに用いることができる。

#### 5. 主な発表論文等

##### 〔雑誌論文〕(計 1 件)

1. Wang Renzhi, Mizuho Iwaihara, "Estimating Reference Scopes of Wikipedia Article Inner-links," Journal of Information Processing, IPSJ, Volume 26, pp. 562-570, July 2018. 情報処理学会論文誌データベース (TOD), 11(2), 9 pages, July 2018(査読有り).

##### 〔学会発表〕(計 8 件)

1. Chen Zihang, Mizuho Iwaihara: Detection of Bursty and Significant Keyphrases from Wikipedia edit history, 6th IEEE International Conference on Big Data and Smart Computing (BigComp 2019), Kyoto, pp.99-103, Feb. 2019(査読有り).
2. Qianwen Wang, Mizuho Iwaihara: Deep Neural Architectures for Joint Named Entity Recognition and Disambiguation, 6th IEEE International Conference on Big Data and Smart Computing (BigComp2019), Kyoto, pp.152-105, Feb. 2019(査読有り).
3. Li Junhao, Mizuho Iwaihara: Two-encoder pointer-generator network for summarizing segments of long articles, DEIM Forum B4-2, Sasebo, March 2019(査読無し).
4. Yuming Guo, Mizuho Iwaihara: Selecting Article Segment Titles Based on Keyphrase Features and Semantic Relatedness, ESKM 2018: 9th International Conference on E-Service and Knowledge Management, Yonago shi, Japan, July 2018(査読有り).
5. Linfeng Yu, Mizuho Iwaihara: Finding high quality documents through link and clickgraphs, 7th International Congress on Advanced Applied Informatics: ESKM 2018: 9th International Conference on E-Service and Knowledge Management, Yonago shi, Japan, July 2018(査読有り).
6. Xu Zhuoyang, Mizuho Iwaihara: Fast Identification of Topic Burst Patterns Based on Temporal Clustering, SCAI 2018: 6th International Conference on Smart Computing and Artificial Intelligence, Yonago, Japan, July 2018(査読有り).
7. Shuangshuang Cai, Mizuho Iwaihara: Joint Embedding of Hierarchical Structure and Context for Entity Disambiguation, SNPD 2017, Ishikawa, Japan, June 2017(査読有り).
8. Liu Yixuan, Gao Zihao, Mizuho Iwaihara: Identifying Topic Evolutionary Patterns Based on Bursty Phrase Clustering, APWeb-WAIM 2017, Beijing, China, July 2017(査読有り).

##### 〔図書〕(計 0 件)

##### 〔産業財産権〕

出願状況 (計 0 件)

名称:

発明者:

権利者:

種類：  
番号：  
出願年：  
国内外の別：

取得状況（計 件）

名称：  
発明者：  
権利者：  
種類：  
番号：  
取得年：  
国内外の別：

〔その他〕  
ホームページ等

## 6. 研究組織

### (1) 研究分担者

研究分担者氏名：  
ローマ字氏名：  
所属研究機関名：  
部局名：  
職名：  
研究者番号（8桁）：

### (2) 研究協力者

研究協力者氏名：  
ローマ字氏名：

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、研究者個人に帰属されます。