

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 17 日現在

機関番号：32689

研究種目：基盤研究(C) (一般)

研究期間：2013～2015

課題番号：25330367

研究課題名(和文) ソーシャルメディアにおける派生過程マイニングとプライバシーマネジメント

研究課題名(英文) Mining derivation histories and privacy management on social media

研究代表者

岩井原 瑞穂 (Iwaihara, Mizuho)

早稲田大学・理工学術院・教授

研究者番号：40253538

交付決定額(研究期間全体)：(直接経費) 3,800,000円

研究成果の概要(和文)：WikipediaやFacebookなどのソーシャルメディアを対象として、ユーザの投稿行動により生じる履歴や、投稿行動の動機分析を主な目的とする。Wikipediaの編集履歴において、バージョン集合の派生や手戻りをバージョン間に含まれる共通文字列を用いて、正確に再構築するアルゴリズムを開発した。またユーザの投稿行動について、投稿の動機を質問票により調査し、新たな友人を獲得する動機、および既存の友人と交流する動機などのモデルを構築した。さらにユーザプロフィールや投稿量などの指標から、投稿動機を高精度で推測できることを示した。

研究成果の概要(英文)：We investigate analysis of contents posted by users of SNSs, and users' motivations on SNS posting. From Wikipedia edit histories, we developed an efficient and accurate algorithm to reconstruct version histories, which can indicate how branches of versions were happened, for illustrating the process of the evolution of articles. We asked Facebook users about their motivations on their postings, and developed a purpose model that can classify into major purposes, including interaction between new online friends and interaction with known friends, and then discovered predictors of these purposes from users' profiles and metrics on posting behaviors.

研究分野：社会情報学

キーワード：ソーシャルメディア データマイニング プライバシー アルゴリズム ウェブマイニング

1. 研究開始当初の背景

Wikipedia や Facebook に代表されるソーシャルメディアは大規模かつ高速な情報共有を行うインフラを通して、多数のユーザがコンテンツやコメントの投稿を行なう双方向のユーザ参加型メディアである。本研究はソーシャルメディアの先導的応用の観点から技術開発を行う。

2. 研究の目的

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。研究テーマのひとつとして、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、その応用展開を行う。

(1) ソーシャルコンテンツの派生過程マイニング

ユーザのコンテンツ編集において、派生によりバージョン集合が生じているコンテンツの派生関係を正確に再構築し、派生時期の推定や最新性・成熟度の評価等への応用を行う。

(2) ソーシャルメディアにおけるプライバシーマネジメント

ユーザの目的にあったプライバシー情報のマネジメントという観点から、大規模なマイニングを用いて適切なプライバシーポリシーを推薦する手法の開発を行う。

3. 研究の方法

代表的なユーザ参加型ソーシャルメディアである Wikipedia には、百科事典の項目として投稿された膨大な記事に加え、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。しかし Wikipedia の編集履歴は、バージョンの投稿者と日付、注釈等の基本情報のみからなり、編集者は差分を繰り返し求めて依存関係を把握する必要があった。バージョンが直前のものから作成されたのか、あるいは手戻りや統合が起きたのかといった依存関係・意味的關係を表すバージョングラフの再構成が必要であるが、(a)従来の文書の差分アルゴリズムは2つの文書間であるためバージョングラフの再構成には計算コストがかかりすぎることに、(b)単に編集距離が最短のものを親バージョンとしても正しい結果が得られず、また人間にとって理解の難しい不自然な差分を生成するという問題点がある。これまで編集履歴の可視化や、編集過程での協調行動の分析、編集回数等を用いた記事の品質評価などの研究があるが、いずれも正確なバージョングラフに基づいていない。

一方、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。

4. 研究成果

(1) バージョングラフの再構築アルゴリズム

バージョングラフを正確に再構築し、さらに記事の分割や統合といった抽象度の高い依存関係を推定する方法を開発した。バージョンの範囲において変更されていない極大の部分文字列を *supergram* と呼び、*supergram* に基づいてバージョングラフを構築する方法を提案した。評価実験により既存の手法よりも提案手法で良い精度が得られることが分かった。さらにバージョングラフを可視化し、共同編集作業の状況を容易に把握できる応用システムの開発を行った (APWeb14 の best demo を受賞)。

(2) フレーズラベリング手法

記事中のパラグラフなどの構造単位を代表する重要なフレーズを抽出することは、索引構築や要約といった応用に加え、知識の構造化という観点からも重要である。フレーズマイニングの手法として、Jiawei Han らの FP-growth をベースにしたものが知られているが、タイトルやアブストラクトといった比較的短い文章を対象にしている短所がある。本テーマでは Wikipedia の記事のような、長文や多数のパラグラフを持つ文章を対象とした **フレーズラベリング手法** を開発した。全く同じフレーズが複数個所に出現するとは考えにくく、編集履歴において派生関係にあるフレーズ集合、あるいは記事間で同一のトピックを述べているフレーズ集合の中から、代表するフレーズを選ぶ戦略を採用し、汎用性の高いフレーズを抽出して知識の構造化に役立てる。有用なフレーズの評価関数を洗い出し、それらの組み合わせにより、最適なフレーズを生成する手法を設計した。もともと記事に与えられている節タイトルに対し、意味的な近さでフレーズの品質を評価する方法

を開発した。

(3) 実体リンキング

固有名詞(人名・地名等)を抽出し、対応する実体を記述する記事等に、曖昧性を解消した上でリンクする実体リンキング問題について、近年はデータグラフを利用して精度を向上する研究が活発に行われている。また関連した問題に一般名詞やハッシュタグの曖昧性解消問題があり、Wikipedia の曖昧性回避ページを用いた多義性の解消方法を発表している (APWeb15)。

(4) Wikipedia リストの要素帰属問題

Wikipedia には同じテーマの記事を列挙したリストが記事として登録されており、例えば東京の高層ビル、日本の超高層建築物、東京のタワー、日本の超高層建築物といった多様で類似した複数のリストが存在する。与えられた記事が含まれるべき最も適切なリストを判定する問題の解法を求めた。

(5) ソーシャルメディアの参加動機モデルとその推定手法

ソーシャルメディアにおいて既知の友人との交流する動機(内向的動機)、ソーシャルメディアであらたな友人を得る動機(外向的動機)、内向的・外向的とは関連のない動機(中立的動機)の3つからなるユーザの動機モデルを提案した。そしてユーザのプロファイル属性(性別、年齢、誕生日等)、情報公開度(顔写真、居住地域、学歴、職業等の公開の有無)、活動度(友人数、投稿写真の数、音楽、映画、本などへのリンク数等)からなる外部から観測可能な因子がこれらの動機と関連するという仮説を設定した。Facebook のユーザ約 250 名へのアンケート調査結果に対する統計的分析で検証を行い、3つの動機それぞれと有意に関連のある観測可能因子を同定し、F1 スコアが最高 0.78 となる予測モデルが構築でき

ることを報告した(J.Computers in Human Behavior) .

(6) 公開コンテンツの投稿意図モデル

投稿を公開する意図の分類モデルとして、何らかの宣伝意図があるか、ある場合はミュージシャンや商品販売といったビジネス宣伝があるか、あるいは単に自身の生活や意見の宣伝を意図しているという分類が考えられる。また宣伝意図のない場合はニュースや他者の投稿を拡散するという意図が挙げられる。このよう投稿公開の意図分類のモデルに対し、予測因子となり得る、外部から観測可能な項目を、Facebook の公開グループでの投稿の分析を通して同定を行った。

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 2 件)

(1) Jianmin Wu and Mizuho Iwaihara, “Revision Graph Extraction in Wikipedia Based on Supergram Decomposition and Sliding Update,” Trans. IEICE, Vol.E97-D, No.4, pp.770-778, April 2014 (査読有り)

(2) Basilisa Mvungi and Mizuho Iwaihara, “Associations between Privacy, Risk Awareness, and Interactive Motivations of Social Networking Service Users, and Motivation Prediction from Observable Features,” Computers in Human Behavior (2015), Elsevier, Vol. 44, pp. 20-34, March 2015 (査読有り)

〔学会発表〕(計 7 件)

(1) Basilisa Mvungi and Mizuho Iwaihara, “Factors that Influence Social Networking Service Private Information Disclosure at Diverse Openness and Scopes,” Proc. SocInfo2013, LNCS 8238, pp. 119-128, Kyoto, Nov. 2013.

(2) Renzhi Wang, Jianmin Wu, Mizuho Iwaihara, “Finding Co-occurring Topics in Wikipedia Article Segments,” Proc. Int. Conf. Asian Digital Libraries (ICADL

2014), LNCS 8839, pp. 252-259, Chiang Mai, Sep. 2014.

(3) Jianmin Wu, Mizuho Iwaihara, “WikiReviz: An Edit History Visualization for Wiki Systems.” Proc. Asia-Pacific Web Conf. (APWeb2014), LNCS8709, pp. 670-673, Changsha, Sep. 2014 (best demo award).

(4) Bonan Li, Jianmin Wu, and Mizuho Iwaihara, “Tracking Topics on Revision Graphs of Wikipedia Edit History,” Proc. Web-Age Information Management (WAIM 2014), LNCS 8485, pp. 204-207, Macau, June 2014.

(5) Qian Chen and Mizuho Iwaihara, “Iterative Algorithm for Inferring Entity Types from Enumerative Descriptions,” Proc. 4th Workshop on Data Extraction and Object Search (DEOS2014), WWW’14 Companion, pp.1285-1290, Seoul, April, 2014.

(6) Wang Mengmeng, Mizuho Iwaihara, “Hashtag Sense Induction Based on Co-Occurrence Graphs,” Proc. 17th Asia-Pacific Web Conference (APWeb 2015), Lecture Note in Computer Science (LNCS) Vol. 9313, pp.154-165, Guangzhou, Sep.2015.

(7) Shan Liu, Mizuho Iwaihara, “Extracting Representative Phrases from Wikipedia Article Sections,” DEIM Forum C3-6, Fukuoka, March 2016.

(8) Basilisa Mvungi, Mizuho Iwaihara, “Estimating Purposes of Users in Social Networking Service Public Contents,” DEIM Forum A5-2, Fukuoka, Japan, 2016.

〔図書〕(計 件)

6. 研究組織

(1) 研究代表者

岩井原 瑞穂 (Mizuho Iwaihara)
早稲田大学・理工学術院・教授
研究者番号： 40253538

(2) 研究分担者

()
研究者番号：

