

科学研究費助成事業 研究成果報告書

平成 29 年 5 月 7 日現在

機関番号：32689

研究種目：基盤研究(B)（一般）

研究期間：2013～2016

課題番号：25280113

研究課題名（和文）Web情報を対象とした数十万人規模での著者推定

研究課題名（英文）Authorship Identification for Hundred-thousand-scale Microblog Users in the Web

研究代表者

山名 早人（YAMANA, Hayato）

早稲田大学・理工学術院・教授

研究者番号：40230502

交付決定額（研究期間全体）：（直接経費） 10,400,000円

研究成果の概要（和文）：インターネット上には様々な情報が氾濫し、その信憑性が社会問題化してきている。本研究では、信憑性を判断するための方法の一つとして著者推定技術をSNS等の短い文書に適用し10万人の候補者の中から、著者を発見する研究に取り組んだ。つまり、事前に著者の書いた何らかの文書があれば、発信者を推定することが可能となる。結果、10万人のSNSユーザの中から特定のユーザを発見するために、30発言あれば60%の精度で発見できる仕組みを構築した。また、上位10位までに抽出できる確率は74%を達成した。これは、世界の他の研究が10万人を対象として精度20%程度に留まっているのに対し大きな貢献である。

研究成果の概要（英文）：Since various information floods on the Internet, its credibility is becoming a social problem. In this study, we researched on authorship identification technique for short messages such as SNS, targeting to identify the authorship of the messages from among 100,000 candidates. That is, if there is some documents written in advance by the author, it is possible to estimate the writer. As a result, we have established a mechanism to find a specific user out of 100,000 SNS users with accuracy of 60% if we have only 30 messages. In addition, the probability of being able to extract in the top 10 places was 74%. This is a major contribution to the fact that other research in the world is limited to about 20% accuracy for 100,000 candidates.

研究分野：ビッグデータ解析

キーワード：著者推定 インターネットの安全性 SNS tweet 専門性推定 信憑性

1. 研究開始当初の背景

(1) インターネット上には様々な情報が氾濫し、その信憑性が社会問題化してきている。信憑性を判断するための方法の一つとして、「同一の事象が記述されている Web 上の記事数」の多数決により判断する方法がこれまで提案されている。しかし、誤った情報が多数を占める場合もありうまく機能しない。

(2) 一方で文書の執筆者を推定することができれば、インターネット上の情報の信憑性を判断する一つの重要な材料となり得る。しかし、従来の著者推定技術（文書中の単語出現頻度分布の違いにより著者を推定する技術）は、小説などの同一話題の文章を対象にしてきており、そのままでは利用することができない。

(3) 異なる話題にも頑健な著者推定手法の研究が行われている。しかし、従来と同様、単語の出現頻度を基本的な特徴量として用いていることから、話題が変化した場合に精度向上ができない。具体的には、1,000 人から特定著者を推定する場合の精度は 60%程度に留まり、インターネット上の膨大なデータを対象に適用することができない。つまり 10 万人等の大規模利用者を対象とした場合、精度が極めて低くなり実用的ではない。

(4) 昨今利用が急速に拡大しているソーシャルネットワークワーキングサービス(SNS)では、一つの文書（記事）が短く、従来の長文を対象とした著者推定の利用ができない。このため、短い記事に対応した著者推定が必要となってきた。

2. 研究の目的

(1) 本研究では、「誰がその記事を書いたか」という著者推定を数十万人規模で行う技術を研究開発することにより、将来的には、記事の著者を推測し、著者本人の信頼性に帰着させることを目指す。具体的には、インターネット上の情報（特に SNS）に対する信憑性判断のための新しい基準（著者判定のための特徴量）を提供し、以て、安心できるインターネット社会に寄与することを目指す。

(2) 本研究は、インターネット上の情報に対する信憑性判断を目的として著者推定を応用することを考えており、話題が異なる文章が記述されたとしても同一著者であるという判定が行える手法を数十万人規模の著者に適用できるようにすることを目指す。

(3) インターネットの安全性という観点から、著者推定に留まらず、幅広い応用へ発展させるための応用研究を合わせて実施し、技術的な網羅性向上を目指す。

3. 研究の方法

(1) インターネット上の様々なデータに対応した著者推定を可能とするため、主に従来手法が苦手とした短文を対象とした著者推定手法の提案を行う。提案にあたっては、従来手法の問題である (a) 多種多様な話題に起因する精度低下、(b) 推定対象となる文章の短文化による特徴量選択の失敗、(c) 膨大な候補者群による計算時間の増大を解決する。評価においては、10 万人規模の Twitter データを用いた実践的な評価を行う。

(2) 著者推定精度を向上させるために「著者のプロフィール推定技術」「著者が用いている単語群から特定分野への専門性を判定する技術」「インターネット上への記事を掲載する際に取得可能な様々な特徴量（アクティブ認証技術）」について研究を進める。

4. 研究成果

(1) 多種多様かつ短い文章から著者の特徴量を抽出するための特徴量抽出手法を提案し、10 万人の Twitter ユーザを用いた評価（30tweet による著者推定）において、MRR（平均逆順位）で 0.65、上位 1 位の精度 (P@1) で 0.60、上位 10 位までの精度 (P@10) で 0.74 を達成した（表 1）。これは、正解ユーザを平均して 1.54 位以内にランキングできることを示しており、上位 1 位の候補が正解である確率が 60%であり、上位 10 位以内に正解ユーザが入る確率が 74%であることを示している。この結果は、従来研究の P@1 で 0.20（Narayanan らの手法で 10 万人に対する評価結果）に比較して大幅な精度向上を示しており、実用化に資する結果を得ることができた。

表 1 提案手法の評価結果（10 万人）

		MRR	P@1	P@10	P@100
実験回数	1	0.668	0.623	0.752	0.856
	2	0.622	0.573	0.719	0.859
	3	0.637	0.585	0.728	0.858
	4	0.644	0.593	0.733	0.866
	5	0.657	0.608	0.746	0.846
Average		0.646	0.596	0.736	0.857

(2) 短い文章を対象とした著者推定を実現する手法として、文章を文字単位で分割し 1 文字、2 文字、3 文字（それぞれ 1-gram, 2-gram, 3-gram と呼ぶ）の出現頻度を基本的な特徴量として採用した。これは、従来提案されている様々な特徴量（品詞、単語、文字）を網羅的に調査した結果、最も精度を出すことができる特徴量として選定したものである。さらに、この方法だけでは十分な精度を達成できなかったことから、n-gram の特徴量に n に比例して大きな重みを与える手法を考案した。図 1 に示しているのは、twitter ユーザ数 1000 名時の評価であり、n に比例して重みを付与することで、大きな精

度向上が得られることを示している (MRR が 0.539 から 0.796 に改善)。

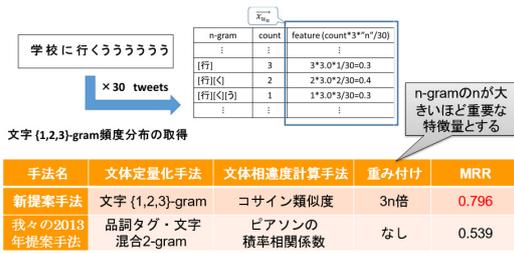


図1 文字{1,2,3}-gram 利用と重み付け

(2) 多種多様な話題に頑健な著者推定を実現するために、1人のユーザは複数の話題を持つことを前提に学習用データセットを構築する手法を提案した。具体的には、候補者となるユーザ (10万人) の投稿したツイート全てを用いて、ユーザ毎に時間軸上で発信時間が異なる k 個の学習データセットを作成した。これは、1ユーザが k 個の異なる話題を持つことを想定している。話題分類においては、 k -means 法を用いてクラスタリングを行った。図2に示す通り、最近の投稿を含め、古い投稿から合計で k 個の特徴量を取得した。なお、学習データセット作成の際、(1)で述べた n -gram の n に応じた重み付けを行っている。高速化に関しても、この k の値を調整する (小さくする) ことで高速化が可能となるが、一方で精度は低下する。また、高速化については、Locality-Sensitive Hashing を用いる手法についても提案し、精度を適用前比で 6% 低下させることを許せば約 2 倍の高速化が可能であることを示した。

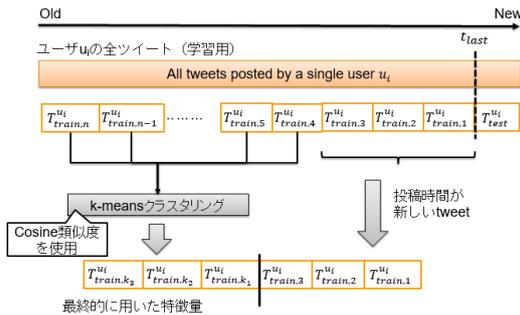


図2 時系列上で複数特徴量取得

(3) テストデータセット (検索対象ユーザの tweet) と学習用データセットとの比較においては、従来手法が用いている SVM 等の機械学習を用いず、コサイン類似度を採用した。これは、従来の研究結果から、機械学習が短い文章に対しては適切に動作しないこと分かっているためである。具体的には、図3に示すように、作成した学習データセットとテストデータセットの間で文体相違度 $Dissim_{Top}$ の計算を行い、それらの文体相違度 $Dissim_{Top}$ を昇順に並べ替えることで、ランキ

ング作成を行った。このランキングのうち、推定対象著者が何位以内に含まれているかどうかを用いて著者推定の評価を行った。

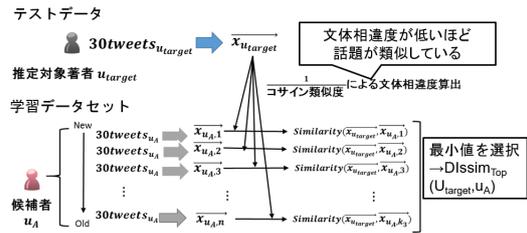


図3 学習・テストデータセット間の比較

(4) 著者の属性 (プロフィール) 推定を行う手法として、他ユーザのプロフィール文から複数抽出する仕組みを構築した。本手法を用いることで、自身がプロフィールを公開していなくても当該ユーザのプロフィール推定が可能となる。図4に示すように、本手法の出力は、推定対象ユーザの属性に関連すると推定された単語のリストである。リスト中の単語は、推定対象ユーザとそれぞれの単語との関連度を示す「単語関連度スコア」によって順位付けされる。具体的には、推定対象ユーザとより深い関係 (ダイレクトメッセージやフォロー、フォロイーの関係) をもつと考えられるユーザのプロフィール文に頻繁に出現する単語に対して高いスコアが付与される。ユーザ間の交流関係をエッジとしたユーザグラフに対して Personalized PageRank を適用すると共に、単語 w に対する IDF を用いて算出する。800 万ユーザのデータを用いて実験したところ、平均して 1 ユーザに対し 7.08~7.25 個の属性を抽出することに成功するとともに、平均精度は、上位 20 位までで 0.72~0.77 であり有効性を確認することができた。

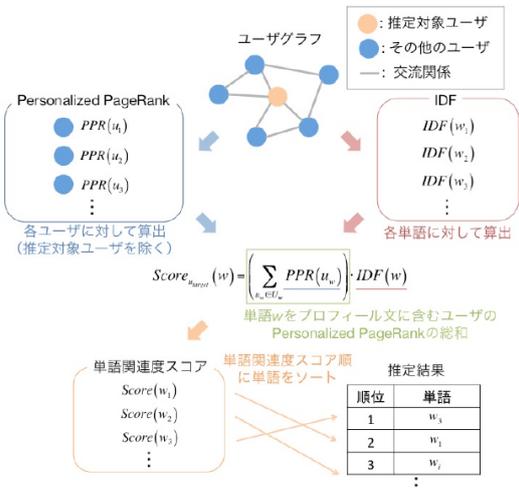


図4 Twitter ユーザ属性推定

(5) 著者の専門性を判定するために、特定分野の専門用語辞書に対して専門度スコアを

付与する手法の提案を行った。特定分野における単語重要度を「一般人が用いない単語であり、かつ特定分野の中でも出現頻度の低い単語がより重要度が高い」という仮説をもって単語重要度を付与する手法を提案した。具体的には、予め専門辞書が与えられている時、当該専門辞書内の単語に重要度を付与する。重要度付与にあたっては、特定分野と特定分野以外のコーパスを用い、特定分野以外のコーパスにはほとんど現せず、かつ特定分野コーパスにも出現頻度の低い単語に高い重要度を付与する。また、ノイズのような意味を持たない単語の重要度が上がるのを防ぐために、文章内での出現頻度に着目し、分散が大きい語句についてノイズと判断する仕組みを構築した。評価では200人のtwitterユーザを対象として、専門分野としてプログラミングを取り上げ、10tweet からどの程度専門性を判断できるかについて実験を行った。正解ランキング（専門性の高いユーザ順に並べる）に対する順位相関係数として0.72を得ることができ、同分野の他の手法に比較して最も高い性能を確認することができた。

(6) 昨今、SNSの投稿ではほとんどの場合、スマートフォン等の携帯端末が利用される。ここでは、今後の応用の可能性を考え、スマートフォン等の携帯端末から情報を入力する際に、実際に本人が利用しているかどうかを自動判別するアクティブ認証の新しい仕組みについても提案、評価を行った。図5に提案システムを示す。まず訓練データ（スワイプ動作）からオンライン分類器を構築する（図中①）。そして、分類器の結果を一定間隔（スライディング Window 方式）で投票し（図中②）、それらの結果を決選投票して（図中③）最終結果とする。最終結果の判定で偽者と判断した場合、2段階目の認証として顔認証などの高度な認証を実施することを前提とする。2段階目の認証の結果、1段階目の判定が間違っていた場合に限り、オンライン分類器を更新する。本提案の貢献は、スライディング Window 方式で認証を実行するた

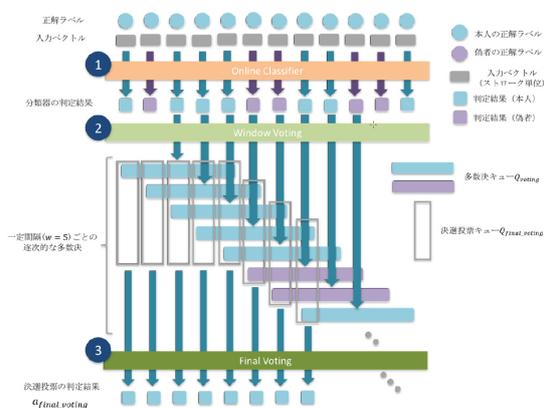


図5 携帯端末を対象のアクティブ認証

め、誤判定が小さくなるという点である。実際、ERR（等価エラー率：誤って受け入れる確率と誤って拒否する確率が同じになる点におけるエラー率）として1.9%（従来手法は6.1%）を達成した。

5. 主な発表論文等

〔雑誌論文〕（計 7 件）

- ① 滝川真弘, 山名早人, 特定分野を対象とした単語重要度計算手法の提案と Twitter における専門性推定への適応, 第 15 回情報科学フォーラム論文集, Vol. 2, No. RD-001, 査読有, 2016, pp. 1-7
- ② 山名早人, ビッグデータ関連プログラム—米国と EU における動向—, 情報処理, Vol. 56, No. 10, 査読有, 2015, pp. 962-967
- ③ Lan Wang, Masahiro Tanaka, Hayato Yamana, What is your Mother Tongue?: Improving Chinese Native Language Identification by Cleaning Noisy Data and Adopting BM25, Proc. of IEEE International Conference on Big Data Analysis, 査読有, 2016, pp. 1-6, DOI:10.1109/ICBDA.2016.7509793
- ④ Kazuya Uesato, Hiroki Asai, Hayato Yamana, Predicting Various Types of User Attributes in Twitter by Using Personalized PageRank, Proc. of IEEE International Conference on Big Data 2015, 査読有, 2015, pp. 2825-2827, DOI:10.1109/BigData.2015.7364090
- ⑤ Syunya Okuno, Hiroki Asai, Hayato Yamana, A Challenge of Authorship Identification for Ten-thousand-scale Microblog Users, Proc. of IEEE International Conference on Big Data 2014, 査読有, 2014, pp. 52-54, DOI:10.1109/BigData.2014.7004491
- ⑥ 奥谷貴志, 山名早人, メンション情報を利用した Twitter ユーザプロフィール推定, DBSJ Journal, Vol. 13-J, No. 1, 査読有, 2014, pp. 1-6, <http://dbsj.org/journal/>
- ⑦ 奥野峻弥, 浅井洋樹, 山名早人, マイクロブログを対象とした 5,000 人レベルでの著者推定手法の提案, WebDB フォーラム 2013 論文集, 査読有, 2013, pp. 1-8

〔学会発表〕（計 14 件）

- ① 石山雄大, 山名早人, オンライン学習を用いたアクティブ認証の実現—スマートフォンを対象として—, 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, I5-5, 2017 年 3 月 6 日—8 日, 岐阜県高山市
- ② 篠原正太, 山名早人, ニューラルネットワークを用いた系列ラベリングによる単語分割手法, 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, F7-1, 2017 年 3 月 6 日—8 日, 岐阜県高山市
- ③ Lan Wang, Hayato Yamana, Robust Chinese Native Language Identification with skip-gram, 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, E6-1, 2016年2月29日-3月2日, 福岡県福岡市

④田中博己, 石山雄大, 上里和也, 山名早人, Locality-Sensitive Hashingを用いた大規模な著者推定の高速化, 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, E6-3, 2016年2月29日-3月2日, 福岡県福岡市

⑤石山雄大, 韓正圭, 山名早人, Twitter アクティブ認証精度向上のための文字 N-gram IDF の提案, 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, F5-2, 2016年2月29日-3月2日, 福岡県福岡市

⑥上里和也, 浅井洋樹, 山名早人, Personalized PageRank を利用した網羅的 Twitter ユーザ属性推定, 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, D2-2, 2016年2月29日-3月2日, 福岡県福岡市

⑦篠原正太, 上里和也, 山名早人, マイクロブログにおける単語間の依存性を考慮した語義曖昧性解消, 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, G8-4, 2015年3月2日-4日, 福島県郡山市

⑧上里和也, 浅井洋樹, 奥野峻弥, 山名早人, Twitter ユーザを対象とした属性推定の精度向上-周辺ユーザの属性補完を利用して-, 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, D8-5, 2015年3月2日-4日, 福島県郡山市

⑨奥野峻弥, 浅井洋樹, 山名早人, マイクロブログを対象とした100,000人レベルでの著者推定手法の提案, 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, D8-1, 2015年3月2日-4日, 福島県郡山市

⑩上里和也, 田中正浩, 浅井洋樹, 山名早人, メンション情報を利用した Twitter プロフィール推定における単語重要度算出手法の考察, 情報処理学会データベース研究会第159回研究発表会, 2014-DBS-159, No. 22, 2014年8月1日-2日, 福岡県福岡市・九州大学

⑪奥野峻弥, 浅井洋樹, 山名早人, マイクロブログを対象とした著者推定手法の提案-10,000人レベルでの著者推定-, 情報処理学会データベース研究会第159回研究発表会, 2014-DBS-159, No. 12, 2014年8月1日-2日, 福岡県福岡市・九州大学

⑫奥谷貴志, 山名早人, メンション情報を利用した Twitter ユーザプロフィール推定, 第6回データ工学と情報マネジメントに関するフォーラム, B2-3, 2014年3月3日-5日, 兵庫県淡路島

⑬上里和也, 奥谷貴志, 浅井洋樹, 奥野峻弥, 田中正浩, 山名早人, twitter におけるアカウント乗っ取りによる成りすましツイートの検出手法の提案, 情報処理学会データベース研究会第158回研究発表会, 2013-DBS-158, No. 21, pp. 1-6, 2013年11月26日, 京都府・京都大学

⑭奥野峻弥, 浅井洋樹, 山名早人, マイクロブログを対象とした1000人レベルでの著者推定手法構築に向けて, 情報処理学会データベース研究会第157回研究発表会, 2013-DBS-157, No. 7, pp. 1-6, 2013年7月22日-23日, 北海道大学学術交流会館

6. 研究組織

(1) 研究代表者

山名 早人 (YAMANA, Hayato)
早稲田大学・理工学術院・教授
研究者番号: 40230502

(2) 連携研究者

大山 敬三 (OYAMA, Keizo)
国立情報学研究所・教授
研究者番号: 90177022

宇野 毅明 (UNO, Takeaki)

国立情報学研究所・教授
研究者番号: 00302977

(3) 研究協力者

奥野 峻弥 (OKUNO, Syunya)
奥谷 貴志 (OKUTANI, Takashi)
浅井 洋樹 (ASAI, Hiroki)
上里 和也 (UESATO, Kazuya)
田中 正浩 (TANAKA, Masahiro)
篠原 正太 (SHINOHARA, Shota)
石山 雄大 (ISHIYAMA, Takehiro)
Lan Wang