様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

# 科学研究費助成事業　　研究成果報告書

平成 ３１ 年　 ４ 月 ２６ 日現在

機関番号：１２６０１
研究種目：挑戦的萌芽研究
研究期間：2016～2018
課題番号：１６Ｋ１２５４２

研究課題名（和文）オンラインを介して 「前読書家」の読書を触発する方式・環境の開発

研究課題名（英文）Developing methods and environments to stimulate pre-readers' reading habit by using SNS

研究代表者
影浦　峡（KAGEURA, KYO）

東京大学・大学院情報学環・学際情報学府・教授

研究者番号：００２１１１５２

交付決定額（研究期間全体）：（直接経費）　　2,600,000円

研究成果の概要（和文）：図書に言及するツイートに関しては、大規模な書名データベースを用いて図書・非図書の分類問題として課題を定義し、自動投稿と非図書を二段階で判別する手法を開発して実用レベルのパフォーマンスを達成した。また比較実験の結果、関連する課題のSOTAと比するパフォーマンスであることを確認した。触発的なツイート（誘引性・非強迫性を備えた図書に言及する）については学習データの構築仕様を定義しデータを二段階で構築し、判定手法を開発した。研究協力者による評価において、触発性に有意な差が見られることを確認した。
システムはプロトタイプを構成しており、2019年度中に試験公開を予定している。

研究成果の学術的意義や社会的意義
街の書店が激減する中、消極的に図書に触れる機会が減っている。オンライン上では積極的に関心を向けた場合には多様な図書関連情報が得られるが、それは既に読書習慣を身につけた人にとって有効ではあるものの、「前読書家」には有効ではない。本課題は、しばしば日常的な文脈で友人や知人・家族等から消極的に得られる図書関連情報が一定程度、人を読書及び図書選択に導くことを踏まえ、それに対応した機構をSNS上で実現しようとするもので、多くの読書関連サイトが読者を想定しているのに対し「人が読むようになることを促す環境」を構築するという点で独創的であり、また、社会的な重要性も高い。

研究成果の概要（英文）：We developed two-step method to identify Tweets that mention Books (TMBs). The method assumes the existence of comprehensive catalogue of book titles and defined the identification task as classification task. Our method achieved practical level precision and the evaluation suggested that our method is comparable to SOTA method in a related task of spam/bot detection.
To identify stimulating tweets (attracting readers and non-imposing), we developed a ML-based method. To facilitate learning, we constructed datasets with a series of tags that indicates the degree of stimulation that TMBs have. An evaluation experiment with over 30 human evaluation participants show that what we judge as "stimulating" tweets are significantly more stimulating than randomly chosen tweets.
Currently we have a prototype system that entice pre-readers into reading, which is under updating. We will make the experimental system available in AY2019.

研究分野：図書館情報学

キーワード：Twitter　感情要因　触発性　データアノテーション　図書に言及するツイート　態度分析　強迫性　誘引性

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９、ＣＫ－１９（共通）

## １．研究開始当初の背景

近年の高度情報化に伴い、スマートフォンをはじめとする個人用情報端末の利用時間が増大し、 電子書籍の普及とともに本が携帯端末の内部で消費されつつある。このことは、将来世代にとって本と無意識に接触する機会が従来世代とは全く異なるものとなりうることを意味する。社会的にも学術的にも広く認識されている読書の教育学的意義を鑑みれば、情報ひいては教育格差に潜在的影響を及ぼすことが懸念される。一般に電子書籍による読書と紙の本との比較は、読書習慣を既に得ている者の視点が暗黙に仮定されてなされがちである。このことは、街の書店の急減による受動的な図書への接触機会の減少といった状況が、電子書籍の状況とともに、そもそも人が本を読むようになることに対してどのような影響を与えるかへの目配りが効きにくいこととつながっている。

## ２．研究の目的

我々は、情報化の展開に即したかたちで、オンラインにおいて読書 を触発できる環境を実装すること、とくにあまり本を読まない状態から読む状態への移行を促す環境 を想定した図書推薦システムの研究が重要となるとの着想に至った。このとき、内容や関心に基づく手法や協調フィルタリングといった、積極的なユーザを前提としてきた従来の図書推薦システム とは異なり、受動的で情報行動が非積極的なユーザを対象とすることが本質的に重要になる。

本研究では、読書意欲はあっても本を読む習慣・本を積極的に求める習慣を持つに至っていない「前読書家」を対象とする。知人・友人による図書への言及が読書行動を誘発しうるとの予備的検討に基づき、SNS の中でも気軽な会話がなされる Twitter を利用し、近しい人との(近接性)日常的な会話(日常性)の中で、さりげなく(非強迫性)、楽しそうに(誘引性)言及される図書の情報を「前読書家」に通知する触発的な図書推薦システムを構築する。そのために必要な概念の整理と要素技術の開発を行う。

## ３．研究の方法

(1) 背景の調査と概念の整理、(2) 要素技術の開発、(3) システムの評価からなる。要素技術には、(2-a) 図書に言及する tweet を同定する技術、(2-b) 触発的な TMB の判定技術がある。

(1) 背景の調査に関してはとりわけ書店の動向、オンライン上の図書推薦システムや書評サイトの動向、電子および紙の出版動向、日本における読書動向などの文献や統計に基づく。概念の整理は、触発性をめぐる概念を心理学等の領域を参照しつつ調査し、データに基づく機械学習問題に落とし込むことが可能な解像度で、触発性に関与する概念を定式化する。
(2) 図書に言及する tweet の同定も触発的な TMB の判定も、機械学習のアプローチで解く。
(3) システムの評価は人間の協力者による実証評価として行う。

## ４．研究成果

### (1) 背景調査と概念整理

関連する研究を整理した結果、「人はどのようにして本を読むようになるのか」に関する研究があまりないこと、それに関して現れた最近の研究が、積極的介入ではなく図書環境への接触の重要性を示していることを確認した。それに対応したオンラインのメカニズムは整備されていないこと、受動的接触の中で触発性を持つメッセージの性質として、特に近接性（近しい人の tweet)、非強迫性（押し付けがましくない）と誘引性（楽しそう・魅力的なかたちで本に言及している）が有効であることを確認した。

### (2-a) 図書に言及するツイートの識別

初年度前半に開発した、書名を含むツイートから TMB を識別する機械学習手法は書名の周りの単語に着目する手法で一定の性能（F 値で 0.68）を達成した。初年度後半に素性設計の改善を行い、また、スパム／ボット投稿的な tweet の同定と書名と同じ単語列ではあるが書名ではないノイズ tweet の識別のステップを分ける 2 段階パイプラインを導入することで、F 値で 0.76 まで性能の向上を実現した。なお、便宜的にここでは F 値を用いるが、実装の際に戦略的な書名言及 tweet の選択においては F 値は必ずしも有効性の指標にはならないことを述べておく。我々は、精度と再現率のトレードオフについても詳細に検討しているが、煩雑になること、またその活用は応用の戦略に依存するため、ここでは述べない。

素性としては、環境に関わるもの（投稿情報等）と tweet に関わるもので異なる効き方をしている。後者については、一般的な語に加え、書名以外の書誌情報の考慮が（当然ではあるが）大きく貢献する。また、「書名らしさ」を具体的な書名とは別に考慮することの有効性も検討した。関連するボット同定や固有名判定の State-of-the-art（SOTA）システムと比較し、既存の Twitter ボット判定アルゴリズム（F 値 0.65）や、固有名抽出アルゴリズム(F 値 0.44)を上回る性能があることを示した。技術的な成果は電子図書館系の国際会議で発表したほか、全体を

まとめた成果は、本報告書作成時に応用情報処理の国際論文誌に ready to accept となっており最終版の投稿と最終判定を待っている状態である。

(2-b) 触発的な TMB の判定機構の開発
触発性に関する属性を付与したデータの構築とそれを学習に用いた判定手法の開発を行った。

データ構築：TMB への触発性アノテーション
　第二年度から第三年度に、誘引性と強迫性、文脈類型（どのような目的で図書に言及しているか）に関する詳細なラベルを付与するための仕様を定義し、学習用データの構築を二段階に分けて行った。第一段階では 10,000 件の tweet に 7 種類のラベルを、第二段階では 9,127 件の tweet に 2 種類のラベルを付与した。これらの一部は、2018 年度 3 月の言語処理学会で発表している。

触発性判定機構の技術開発
　これを用いて、誘引性を評価する機構を開発した。誘引性は単純な感情分析とは違い、ネガティブ感情も図書評価にポジティブなことがある（例：泣ける小説）。しかも構築したデータのラベルがアンバランスで、ネガティブが非常に少ない。これは、いわゆるスタンス検出と極めて近い状況であるため、その領域の最先端の機械学習手法を適用することができる。一方、非強迫性には特徴的な構文パターンが認められる。また、強迫性をめぐる事例は実際にはかなり少ないため、機械学習に向かないと判断した。そこで、構文解析結果をもとにしたパターンマッチで、押し付けがましい構文を検出する手法を採用した。なお、近接性はユーザ同士のインタラクションの活発さを「いいね」や RT の数で判定することにした。
　これらは第二年度末から第三年度にかけて、豪 CSIRO 研究所でのツイートに対するスタンス検出タスクに関する共同研究にもつながった。共同研究では、スタンス・データセットの構築、検出アルゴリズムの実装（SVM、RNN モデル、記憶ネットモデル、Transformer モデル）、実験プログラムの実装、スタンス検出のための学習済み単語分散表現のツイートデータからの構築を行っている。現在、この研究成果は国際会議に投稿中である。
　ここまでの作業は基本的に tweet 一つ一つを対象に行っている。しかしながら、しばしば tweet は連投されることがあるため、そのスパンを考慮するために、Twitter の談話スパン（一連の会話のまとまり）を測定する基礎技術を開発した。トピック分析・時系列情報と組み合わせることで、触発的な tweet 前後から図書に関する情報と誘引性を有する情報を取得する可能性を開いた。

(3) システムの評価
　第一年度に非常に基本的なシステムを実装し予備的な評価を行った。また、図書に言及するツイートを模したメッセージの誘引性・非強迫性が読書への触発性に与える影響をオンラインの質問紙で調査した。対象は 34 名の大学生（文系 25 名・理系 9 名）で、調査の結果、システムから出力される正解メッセージ（図書に言及するツイートのうち、誘引性があり非強迫的なもの）が読書を触発する効果があるとの仮説が、分散分析により有意に支持された。この結果に関しても現在国際会議論文を執筆中である。また、開発した機構を組み込んだシステムは 2019 年度内に稼働開始予定である。

## ５．主な発表論文等

〔雑誌論文〕（計　0　件）
（本報告書作成時点で採択最終段階の論文 1 件）

〔学会発表〕（計　7　件）

矢田竣太郎, 影浦峡. 2019. “図書を推薦するツイートの押し付けがましさに関する特徴の分析.” 言語処理学会第 25 回年次大会発表論文集, 842-845.
Shuntaro Yada, Kazushi Ikeda, Keiichiro Hoashi, and Kyo Kageura. 2017. “A Bootstrap Method for Automatic Rule Acquisition on Emotion Cause Extraction.” In IEEE International Conference on Data Mining Workshops (ICDMW) - Sentiment Elicitation from Natural Text for Information Retrieval and Extraction (SENTIRE), 414-421.
Sunghwan Mac Kim, Kyo Kageura, James McHugh, Surya Nepal, Cecile Paris, Bella Robinson, Ross Sparks, Stephen Wan. 2017. “Twitter Content Eliciting User Engagement: A Case Study on Australian Organisations.” In WWW2017 Poster Session.
Shuntaro Yada, and Kyo Kageura. 2017. “Measuring Discourse Scale of Tweet Sequences: A Case Study of Japanese Twitter Accounts.” In The 19th International Conference on Asia-Pacific Digital Libraries (ICADL), 150-157.

矢田竣太郎，岩井美樹，影浦峡．2017．“日本語書籍タイトルの形式的構造の分析．” 言語処理学会第 23 回年次大会発表論文集，699-702
Shuntaro Yada, and Kyo Kageura. 2016. “Improved Identification of Tweets That Mention Books: Selection of Effective Features.” In The 18th International Conference on Asia-Pacific Digital Libraries (ICADL), 150-156.
矢田竣太郎，影浦峡．2016．“図書に言及するツイートの抽出: 素性・データ量・手法の効果に関する考察．” 電子情報通信学会技術研究報告 言語理解とコミュニケーション (NLC), 29-34.

〔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　0　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：矢田竣太郎
ローマ字氏名：YADA, Shuntaro