

論文の内容の要旨

論文題目 Towards a digital environment for passive exposure to books based on online social media
(SNS を用いた受動的図書遭遇環境の構築に向けて)

氏 名 矢田 竣太郎

本論文では、伝統的に物理環境で担保されてきた受動的図書遭遇 (*passive exposure to books*) をデジタル環境上で提供するシステムの必要性を提起し、ソーシャルネットワーキングサービス (SNS) 上でなされる図書への言及を提示する手法で実現できると提案した。さらに、そのようなシステムの実現に向けて、ユーザを読書へと触発するために必要な要件 (触発性; *inspiringness*) を整理・識別し、実装可能なシステム構成を設計した。本論文では、提案したシステム構成に基づく実装の基盤部分も射程に入れた。すなわち、システムを構成するモジュールのうち自然言語処理 (NLP) を扱うコア NLP モジュール群が既存の NLP タスクとは新規の問題設定を扱うことから、そうした問題設定を NLP のタスクとして定義し解決することを本論文におけるシステム実装の目標に定めた。リサーチクエスチョンはしたがって以下の 3 点に集約される。

- デジタル環境において図書遭遇を実現するために SNS 上の図書への言及をユーザに提示する場合、求められるシステムの要件=触発性は何であるか？
- そのような要件を満たすシステムの実現可能な設計とはどのようなものか？
- システムを構成する各モジュールはどうやって実装することができるか？

そして本論文を通じた貢献は以下の 5 点にある。

- 読書をめぐる環境の現状に基づき、受動的図書遭遇をデジタル環境上に実現することが、情報化に伴って拡大すると予見される潜在的な教育格差の是正に必要であることを示した。
- システムの対象ユーザに含まれる、あまり読まない人々をも読書へと触発するために必要な要件 (触発性) を識別した。
- システムの実現可能な設計と、システムを構成するモジュールを実装するために解かねばならないタスクの定義を行った。
- データセットを構築し、詳細な分析を加えた上で、実用的な水準でコア NLP モジュール群のためのタス

クを解き、エラー分析に基づいて展望を示した。

- 本論文で実装の射程に含めなかったユーザインタフェース・ユーザエクスペリエンス (UI/UX) を扱うモジュール群について、扱うべき問題や実装への道程をまとめた。

以下では、本論文を章ごとに要約する。なお、第1章・第2章・第5章の一部、および第7章の大部分は Yada, Kageura, and Paris (2019) に基づく。

Part I: Background and research questions

■Chapter 1: Introduction 本章では図書遭遇システムが現代において必要となる理由を明らかにした。まず Sikora, Evans, and Kelley (2019) などの鍵となる文献を参照しながら、読書それ自体のみならず家庭の書棚に代表されるような受動的図書遭遇がリテラシーに与える重要性を指摘した。次に、伝統的に物理環境が受動的図書遭遇を担保してきたものの、書店の減少や電子書籍の普及に伴いその機能が低下しており、代わって存在感を増すデジタル環境はそのオンデマンド性と個人化アルゴリズムにより受動的図書遭遇をうまく提供できていない点に言及した。進展する情報化およびあまり読まない人々が特に被る潜在的な教育格差を考慮すると、デジタル環境に受動的図書遭遇が配置されるべきであり、意図せぬ情報遭遇の場となる SNS がその目的に活用できることに触れた。SNS 上でなされる図書への言及を増幅することで、デジタル環境の中でも特に SNS で観測されやすいフィルターバブル (Pariser, 2011) を軽減して、あまり読まない層にも読書に関する話題が届くことが期待できる。このような形で読書を触発するために、図書への言及を提示する仕組みが持つべき性質を触発性と命名し、本論文での検討の対象に据えた。なお、本論文で扱う SNS 上の図書への言及は、日本語による Twitter 上のユーザ投稿 (ツイート) に限定した。

■Chapter 2: Related work 本章では3つの視点から関連研究をまとめた。(1) デジタル環境における読書推進、(2) 図書情報システムおよびタスク、(3) 触発性に関連する概念である。(1) については、受動的図書遭遇機会の減少を考慮して実施される活動やサービスが、明確な読書意欲はあっても読むべき本を決められない人々を対象としていることがわかった。本論文の図書遭遇システムはその一手手前、つまり読書意欲の触発を目的としていることを確認した。(2) には図書情報検索や図書推薦が属するが、いずれも提示すべき図書とユーザの情報需要との“関連性”に最適化することを目標としており、読書家をユーザとする点が本論文のシステムと異なることを確認した。(3) では、目的の対象に積極的な関心を寄せる手前の人々に働きかけるといふ、本論文の触発性と同様の作用を扱う理論や概念として、アフォーダンス、セレンディピティ、態度変容・説得、ロコミ効果を検討し、触発性の具体的構成要素の同定に向けた理論的準備をした。

Part II: Conceptual framework

■Chapter 3: Requirements for the digital exposures to books この章の目的は、触発性の具体的構成要素を同定し、図書遭遇システムに実現可能な設計を与え、本論文で解くべき課題を整理することである。

触発性は以下の4つの構成要素を同定した。

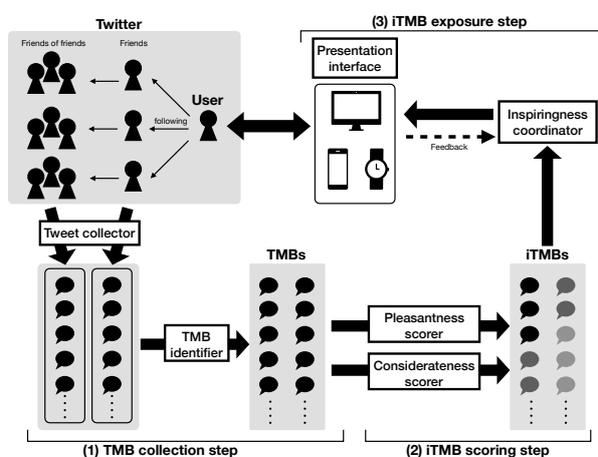


図 1: 図書遭遇システムの設計

日常性 (dailiness): 図書への言及を提示する先のデジタル環境が、ユーザにとってよく利用されていること

近接性 (proximity): 図書への言及が、ユーザにとって任意の尺度で近い人から発信されていること

好評性 (pleasantness): 図書への言及において、発信者が図書を好ましく評価していること

非強迫性 (considerateness): 図書への言及の提示が、ユーザにとって押し付けがましくないこと

システム構成は (1) 図書に言及するツイート (tweet that mention books; TMB) の収集、(2) TMB の触発性測定、(3) 触発的 TMB (inspiring TMB; iTMB) の提示の 3 段階からなるパイプラインとし、計 6 つのモジュールを設計した (図 1)。最後に、本論文で解く課題を、触発性構成要素の必要性検証およびコア NLP モジュール群の実装の 2 つに設定した。

■Chapter 4: Validation of the inspiringness components 触発性を構成する 4 つの要素が実際に、あまり読まない人々を含め、ユーザを読書へと触発することに貢献するか検証するのが本章である。いくつかの要素は関連研究によってその貢献性が確かめられるが、好評性と非強迫性が理論的に衝突するケースがあることから、両者の交互作用を心理実験により検証することとした。TMB を模したメッセージの好評性と非強迫性を調整した質問紙調査の統計的解析により、たとえ好評性があるメッセージでも、押し付けがましく図書を勧める強迫的な言及は読書を触発しにくいことが示され、触発性の構成要素として妥当と結論づけた。

Part III: TMB corpora

■Chapter 5: TMB corpus creation TMB の性質を分析し、コア NLP モジュール群開発に資するために、ラベル付き TMB コーパスを作成した。TMB を比較的多く収集できると期待できる、書名文字列を含むツイートと読書に関連したハッシュタグを含むツイートの 2 種類を準備し、TMB かどうか (2 値) と触発性の中でもテキストに関わる好評性と非強迫性を人手でアノテーションした。本章はそのガイドラインに加え、アノテーション結果としての 2 万件規模の TMB コーパスの概要、および十分高いアノテーション一致度が得られたことを報告している。

■Chapter 6: Descriptive analysis of TMB corpora 2 種類の TMB コーパスのテキスト性質を分析し、コア NLP モジュール群の開発に向けた洞察を得るため、記述統計的な分析を実施した。アノテートされた触発性ラベルの量的分析の結果、好評性のあるツイートが大多数を占めること、一定以上の強迫性が認められる「図書を勧めるツイート」はあまり出現しないことなどが分かった。加えて、図書を勧めるツイートにみられる推薦表現と推薦先対象を質的に分析し、強迫性の強度に応じた言語表現パターンや、典型的に対象とされる集団の性質を整理した。

Part IV: Core NLP modules of the system

■Chapter 7: TMB identification **TMB 識別器**は図書遭遇システムの第 1 ステップを構成する、コア NLP モジュール群の 1 つである。本章において、一般のツイートから TMB を見つける TMB 識別タスクを「書名文字列を含むツイートを TMB・ノイズ・ボットの 3 値に分類するタスク」として定義した。これはまず、本論文で提案した図書遭遇システムが書名文字列を含むツイートを出発点とすることが、得られる TMB の量の多さと性質の偏りの少なさのバランスを考慮した上で妥当だからである。そして、書名文字列を含むツイートには、書名文字列を含むが実際にはその書名の図書には言及していないツイート (ノイズ) と、機械的に生成された大量のツイート (ボット) が含まれ、図書遭遇システムはユーザのデジタル環境上の社交関係に由来する

言及を増幅するのが目的であることから、非 TMB たるノイズとボットはいずれも排除されねばならない。本章では、直截的な 3 値分類器ではなく、ボットフィルターとノイズ除去からなる 2 段階パイプラインを提案し、各種のベースラインを上回る実用的な TMB 識別性能を達成した。

■Chapter 8: iTMB scoring TMB の触発性の測定は図書遭遇システムの第 2 ステップであり、TMB 識別器で得られた TMB のテキストから好評性と非強迫性を測定する 2 つのコア NLP モジュールから構成される。第 6 章の TMB コーパス分析に基づき、好評性測定は図書への評価極性を 3 値（ポジティブ・ネガティブ・ニュートラル）に分類するタスクとして定義した。この定式化により既存の評判分析手法を適用でき、結果として実用的な分類性能を達成した。しかし、極端に少ないネガティブ好評性の分類は改善の余地がある。

非強迫性測定は、強迫的な図書推薦表現の検出タスクとして定義した。第 6 章の質的分析の知見に基づき、推薦表現の強度に応じた表現パターンのルールを設計した結果、TMB の強迫性をおおむね分類できることを示した。構文情報も考慮したルールにすることで、さらなる性能向上が期待できることも指摘した。

Part V: Conclusions

■Chapter 9: Summary 本章では、本論文全体を要約し、Chapter 8 までの各章の要旨をまとめた。

■Chapter 10: Outlooks この章では、本研究の今後の展開を、(i) コア NLP モジュール群の性能向上、(ii) UI/UX 実装、(iii) システム運用の 3 点でまとめた。(i) について、図書遭遇システムがパイプライン構成であり、コア NLP モジュール群はその上流に位置することから、これらの性能向上はシステム全体で取り扱える TMB の品質の高さに直結する。該当する章の末尾にまとめた技術的展望をもとにさらなる改善が期待できる。(ii) に関しては、本論文で扱わなかった UI/UX モジュールの実装の重要性を改めて確認し、著者が試作している実装の 1 例を示した。(iii) は、図書遭遇システムが運用可能になった場合に展開できる研究として、デジタル環境における受動的図書遭遇がユーザを読書へと触発する作用機序の長期・広域的な解明を挙げた。さらに、本論文がコア NLP モジュール群実装のための課題を解いたことで、様々な UI/UX に基づく多様な形態・内容の受動的図書遭遇をデジタル環境に埋め込める基盤を提供した意義に触れた。

文献

Pariser, E., 2011. *The filter bubble: what the internet is hiding from you*. Penguin Press.

Sikora, J., Evans, M.D., and Kelley, J., 2019. Scholarly culture: how books in adolescence enhance adult literacy, numeracy and technology skills in 31 societies. *Social Science Research*, 77, pp.1-15. DOI: 10.1016/j.ssrresearch.2018.10.003.

Yada, S., Kageura, K., and Paris, C., 2019. Identification of tweets that mention books. *International Journal on Digital Libraries*. DOI: 10.1007/s00799-019-00273-4.