

令和 6 年 6 月 25 日現在

機関番号：13601

研究種目：基盤研究(C)（一般）

研究期間：2020～2023

課題番号：20K00800

研究課題名（和文）機械学習による多読図書の難易度推定

研究課題名（英文）Estimating Extensive Reading Text Difficulty Using Machine Learning

研究代表者

Brierley Mark (Brierley, Mark)

信州大学・全学教育センター・外国語准教授

研究者番号：70646877

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：多読支援システムを開発し、10年以上運用し、豊富なデータを蓄積してきました。本研究では、機械学習を用いて多読図書の難易度を自動推定する方法を検討しました。機械学習手法の選定とデータの整形を繰り返し行い、方法の精度を向上させました。実験結果により、この方法の有効性が確認され、多読学習者にとって有用なツールとなることが示唆されました。

研究成果の学術的意義や社会的意義

本研究では、多読図書の難易度を自動推定するシステムの開発に向けた基礎研究を行い、難易度のパラメータを解明しました。学術的には、多読学習の効果を定量的に評価する新しい知見を提供しました。社会的には、学習者が適切な難易度の図書を選択することで、語学力の向上を支援し、教育の質を高める可能性があります。この研究は、教育機関や公共図書館での応用が期待されます。

研究成果の概要（英文）：We have developed an extensive reading support system and accumulated a substantial amount of data over more than 10 years of operation. This study investigated a method for automatically estimating the difficulty of extensive reading materials using machine learning. Utilizing the accumulated data, we iteratively selected machine learning techniques and formatted data to improve the method's accuracy. The experimental results confirmed the effectiveness of this method, suggesting it as a valuable tool for extensive reading learners.

研究分野：Extensive Reading

キーワード：多読

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、研究者個人に帰属します。

様式 C - 19、F - 19 - 1 (共通)

1. 研究開始当初の背景

多読 (Extensive Reading, ER) は、大量の本を比較的流暢に読むことで外国語を習得する方法です。読者がテキストに没頭し、分析するのではなく楽しむことが目標です。この方法は読解力や語学力の向上に非常に有益です。しかし、ER の効果は、学習者が自分の読解力に合った本を選ぶ能力に依存します。ER の成功には、学習者の読解レベルの把握、書籍の読解レベルの把握、学習者に適したレベルの書籍の入手可能性が重要です。

伝統的なテキスト難易度推定方法である Flesch Reading Ease (FRE) 指数などは、日本の学習者のような読者の主観的な経験と常に一致するわけではありません。多読書を出版する出版社は通常、自社の本にレベルを割り当てますが、出版社ごとに異なるシステムやスケールを使用するため、学習者は見慣れない出版社の本のレベルをすぐに理解できないことがあります。日本では、SSS 英語多読研究会によって定義された読解の容易さレベル (Yomiyasusa Level, YL) が、英語学習者のためのテキストの読解可能性を評価するために使用されますが、これは学習者や教師の経験に基づいています。YL データはすべての多読書に対して提供されているわけではなく、出版社は新しいタイトルを次々と出版し、学習者は時折、外国語学習者向けに特別に書かれていない本を読むこともあります。したがって、ER に使用できる資料の難易度をより正確かつ客観的に推定する方法が必要です。

2. 研究の目的

2.1 主目的 (Main Objective)

この研究の主な目的は、ER プログラムで使用される書籍の難易度を自動的に推定するシステムを構築するために必要な基礎的な理解を深めることです。

2.2 具体的な目標 (Specific Goals)

目標は以下の3つです。

- ・ ER 資料の難易度に影響を与えるパラメータを特定して分析する。
- ・ 従来よりも高精度で Yomiyasusa Level (YL) を予測するモデルを開発する。
- ・ 学習者と教育者にとって有効なツールであることを確認するために、実験結果を通じてモデルを検証する。

3. 研究の方法

3.1 データ収集 (Data Collection)

YL に寄与するパラメータを調べるために、テキストの言語的特徴を定量的に評価するウェブベースのソフトウェア Coh-Metrix を使用しました。多読書と一般書のテキストを Coh-Metrix に入力し、11 カテゴリーにわたる 106 パラメータのスコアを計算しました。

3.2 パラメータ分析 (Parameter Analysis)

3.2.1 選択基準 (Selection Criteria)

パラメータが YL と強い相関を持つかどうかを判断するために、集団相関係数の差の検定を行い、YL 予測に有意に寄与しないパラメータを除外しました。

3.2.2 分析ツール (Analysis Tools)

Lasso 回帰分析を使用して、特定されたパラメータに基づいて YL を予測するモデルを開発し、モデルの汎化性能を最適化するためにグリッドサーチ交差検証を使用しました。

3.3 モデル開発 (Model Development)

3.3.1 回帰分析 (Regression Analysis)

Lasso 回帰はデータを正規化し、目的変数に影響を与える説明変数を選択することで過適合を防止します。様々なモデルの順序の決定係数を計算して、YL を予測するためのモデルの順序を決定しました。

3.3.2 交差検証 (Cross-Validation)

グリッドサーチ交差検証を使用して、モデルの汎化性能を評価し改善しました。ハイパーパラメータおよび他のパラメータをグリッドサーチで最高の精度を達成するように設定しました。

4. 研究成果

4.1 主な発見 (Key Findings)

分析により、YLと強い相関を持ついくつかの主要なパラメータ（総単語数、逆説的付加接続詞および比較接続詞の z スコア、因果関係のある接続詞の発生率など）が特定されました。開発された Lasso 回帰モデルは、YL 予測の相関係数が.9 を超え、FRE 指数よりも大幅に高いことが示されました（図 1 および図 2 参照）。

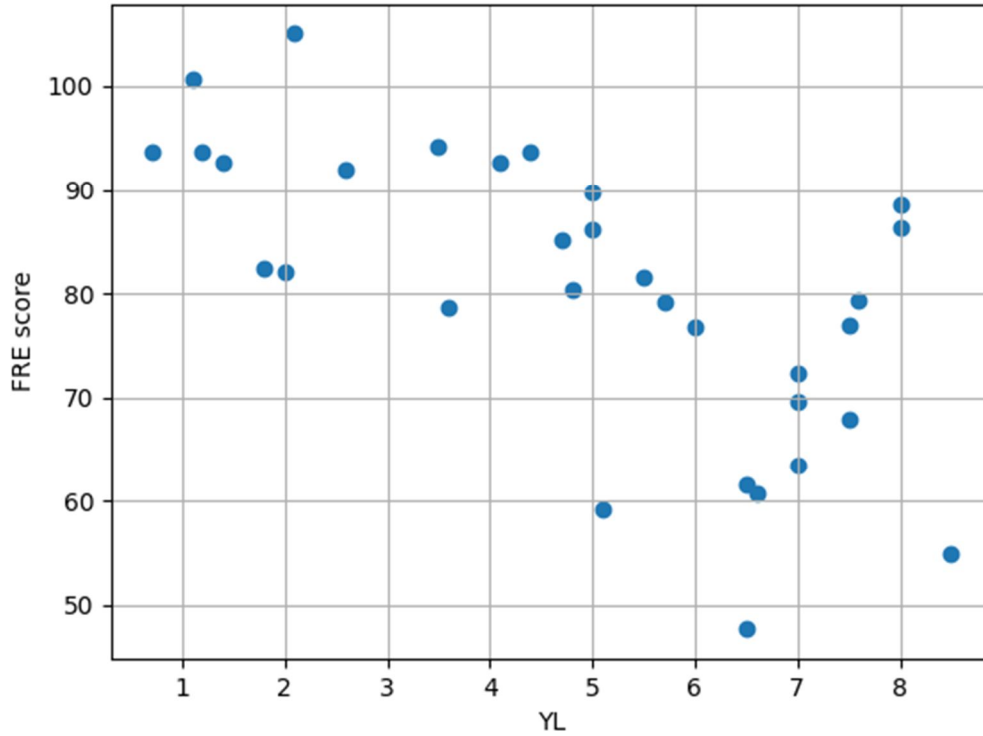


図 1. YL に対する FRE スコア（相関係数 = -0.650）

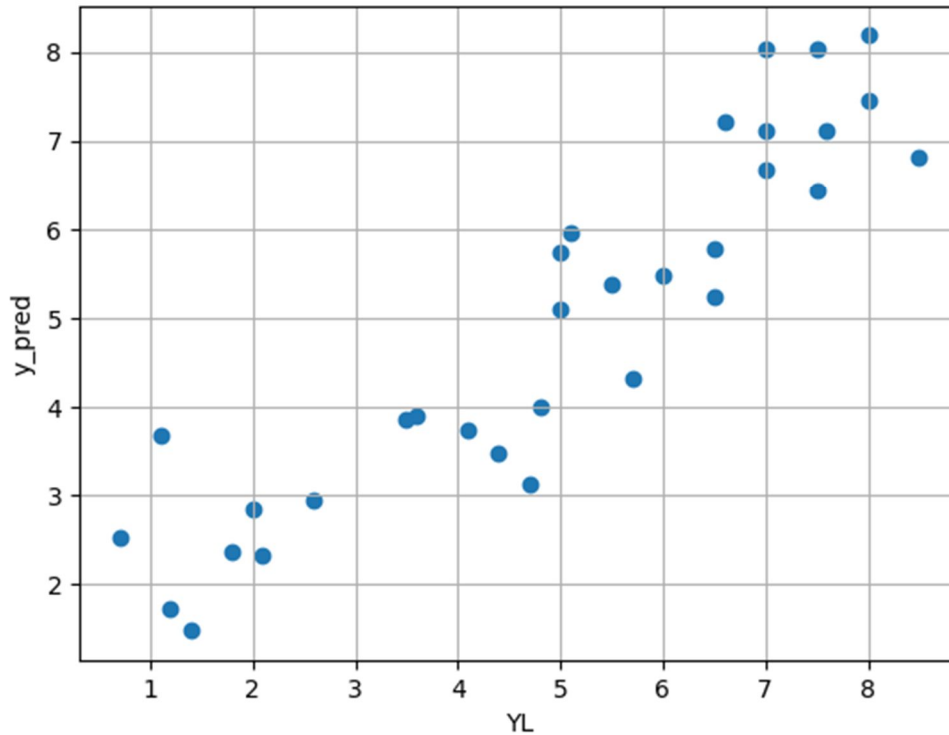


図 2. YL に対する予測 YL (相関係数 = .917)

4.2 学術的意義 (Scientific Significance)

新しい定量的な ER 評価方法が導入され、テキストの読みやすさと難易度に影響を与えるパラメータの理解が深まりました。従来の FRE 指数を超える精度を持つモデルを開発しました。

4.3 社会的意義 (Social Significance)

このアプローチは、学習者が適切なレベルの本を選ぶことを支援することで、語学習得を強化できます。教育機関や公共図書館における ER プログラムの質と効果を向上させ、教育者が生徒のニーズにより適した読書資料を提供するためのツールを提供する可能性があります。

4.4 今後の方向性 (Future Directions)

Coh-Metrix で計算されるパラメータ以外の追加パラメータを考慮することで、モデルの精度をさらに向上させることができます。多様なテキストサンプルを使用したより広範な検証研究により、モデルの予測精度を向上させることができます。

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 2件/うち国際共著 0件/うちオープンアクセス 2件）

1. 著者名 Corinna V. GOTO & Mark BRIERLEY	4. 巻 17
2. 論文標題 Reading The Gruffalo in Another Language: Assessing Non-Graded Children's Picture Books for Extensive Reading	5. 発行年 2023年
3. 雑誌名 信州大学総合人間科学研究	6. 最初と最後の頁 92-104
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Mark Brierley, Ryoma Nakamura, Masaaki Niimura	4. 巻 5
2. 論文標題 Developing an Online Graded Reader Database to Support ER	5. 発行年 2020年
3. 雑誌名 Proceedings of the fifth World Congress of Extensive Reading	6. 最初と最後の頁 219 - 228
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Ruzicka, David	4. 巻 15
2. 論文標題 Teaching Academic English in the Age of AI: Notes on What Translation Software Means for University English Education	5. 発行年 2021年
3. 雑誌名 Shinshu University Journal of Arts and Sciences	6. 最初と最後の頁 133-145
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計11件（うち招待講演 2件/うち国際学会 3件）

1. 発表者名 SHUICHIRO KACHI, MASAAKI NIIMURA, MARK BRIERLEY
2. 発表標題 Using lending history to recommend books for extensive reading
3. 学会等名 JALT PanSIG2022
4. 発表年 2022年

1. 発表者名 SAKAGUCHI RIHO, MASAAKI NIIMURA, MARK BRIERLEY
2. 発表標題 Difficulty estimation method for extensive reading of general english books
3. 学会等名 JALT PanSIG2022
4. 発表年 2022年

1. 発表者名 TOYOSHIMA KOTA, MASAAKI NIIMURA, MARK BRIERLEY
2. 発表標題 Automatic question generation for an extensive reading placement test
3. 学会等名 JALT PanSIG2022
4. 発表年 2022年

1. 発表者名 MARK BRIERLEY
2. 発表標題 The Potential of Machine Translation to Provide Online Extensive Reading
3. 学会等名 JALT2022 全国語学教育学会 第48回年次国際大会（国際学会）
4. 発表年 2022年

1. 発表者名 MARK BRIERLEY
2. 発表標題 Extensive reading ABC: Agency, Books and Counting
3. 学会等名 JALT（全国語学教育学会）岐阜支部（招待講演）
4. 発表年 2022年

1. 発表者名 MARK BRIERLEY
2. 発表標題 Reading your way to better English
3. 学会等名 JALT (全国語学教育学会) 豊橋支部 (招待講演)
4. 発表年 2022年

1. 発表者名 Noda Yuki, Masaaki Niimura, Mark Brierley & Megumi HASEBE
2. 発表標題 A Machine Learning Method for Estimating the Difficulty Level of Graded Readers
3. 学会等名 Extensive Reading Around the World 2021 (国際学会)
4. 発表年 2021年

1. 発表者名 Mark Brierley
2. 発表標題 Testing Extensive Reading
3. 学会等名 Extensive Reading Around the World 2021 (国際学会)
4. 発表年 2021年

1. 発表者名 中村涼真, 新村正明, 増田武浩, Mark Brierley
2. 発表標題 多読支援Webデータベースの開発
3. 学会等名 Extensive Reading Summer Seminar 2020
4. 発表年 2020年

1. 発表者名 Mark Brierley
2. 発表標題 Getting Students Reading: How and Why
3. 学会等名 Yamagata JALT Presentation
4. 発表年 2020年

1. 発表者名 Mark Brierley
2. 発表標題 Online reading e-books in the library
3. 学会等名 Fukuoka JALT Presentation
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分担者	新村 正明 (Niimura Masaaki) (20345755)	信州大学・学術研究院総合人間科学系・教授 (13601)	
研究 分担者	R U Z I C K A D A V I D ・ E D (Ruzicka David) (70436898)	信州大学・学術研究院総合人間科学系・准教授 (13601)	
研究 分担者	長谷部 めぐみ (Hasebe Megumi) (50878725)	信州大学・学術研究院総合人間科学系・准教授 (13601)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------