

書籍レビューの「いいね」予測の説明可能性に基づく ユーザに「刺さる」推薦フレーズの抽出

江田 怜央[†] 牛尼 剛聡^{††}

[†]九州大学芸術工学部 〒815-8540 福岡県福岡市南区塩原 4-9-1

^{††}九州大学大学院芸術工学研究院 〒815-8540 福岡県福岡市南区塩原 4-9-1

E-mail: [†]kouda.reo.109@s.kyushu-u.ac.jp, ^{††}ushiama@design.kyushu-u.ac.jp

あらまし 書籍を購入する際に参考にするものとして、書籍レビューサイトが挙げられる。ユーザは書籍のレビューを読むことで本の内容や評価を知ることができる。その際、ユーザが自身の趣味・嗜好や注目している観点と合致するようなレビューを見つけることができれば、その書籍をさらに読みたくなるようユーザを動機づけることができる。本研究では、ユーザに推薦された書籍に対して、ユーザの興味や関心を惹きつけるようなフレーズをレビューから抽出する手法を提案する。具体的には、書籍レビューサイトにおいてユーザが「いいね」を押したレビューを教師データとして、どのユーザが対象レビューに「いいね」を押すのかを予測する機械学習モデルを構築する。そして、機械学習モデルを解釈する代表的なアルゴリズムの一つである LIME を用いて、対象ユーザが注目すると予想される単語を含むレビュー文を抽出し、ユーザに「刺さる」推薦フレーズとして提示する。

キーワード レビュー, 推薦, 機械学習, 説明可能性

1 はじめに

近年、書籍の評価や感想を投稿する場として、インターネット上での書籍レビューサイトが広く利用されている。例として「読書メーター」¹、「ブックログ」²、「本が好き！」³などが挙げられる。また、オンラインショッピング等で使用する e コマースサイトにおいても書籍のレビューを投稿・閲覧することが可能となっている。ユーザは書籍を購入する際に、それらのサイトに投稿されたレビューを参考にすることができ、レビューを閲覧することで書籍の内容や評価、また読者の感想などの情報を得ることができる。その際、ユーザが自身の趣味・嗜好や注目している観点と合致するようなレビューを見つけることができれば、ユーザはその書籍に対してより強い関心を抱くことができる。

また、書籍レビューサイトでは、ユーザは気に入ったレビューに対して「いいね」を押すことができるものが多い。ユーザは、自身が書籍を読んだ際に感じたことと似た内容がレビューに書かれている場合や、自身の考えとは異なる新たな観点でのレビューに面白さを感じた場合などに「いいね」を押すことが多い。「いいね」を押すことで、自身の考えがそのレビューの内容と近いものであり、そのレビューを肯定的に受け止めているという意志を示すことができる。そのためユーザが「いいね」を押したレビューを分析することで、ユーザの考えや書籍の好みの傾向などを把握できる。これに関連して、Kosinski ら [1] は Facebook を対象とした研究を行い、ユーザの「いいね」の傾

向からユーザの属性が推測可能であることを示した。

さらに近年、機械学習を用いた書籍に限らない様々なアイテムの推薦が広く利用されており、多数のアイテムの中からユーザの好みに合うアイテムを見つけるために必要不可欠な存在となっている。推薦技術の向上によって、ユーザに適した推薦がなされるようになった一方で、アイテムがユーザに推薦された際に「なぜそのアイテムが推薦されたのか」といった推薦理由の説明が適切に提示されていないという問題が存在する。そのため、たとえ推薦システムが良い推薦を行っていたとしても、その推薦理由を知ることができなければ、ユーザはそのアイテムの推薦結果を受け入れることができない可能性がある。ゆえに、ユーザが推薦に納得するために、機械学習モデルの説明可能性を向上させることは重要な課題となっている [2]。現在、「説明可能な AI」と呼ばれる、AI モデルの振る舞いを人間が理解できるように解釈する様々なアルゴリズムが開発されており、それらを用いて推薦結果に対する理由づけを行う研究が進められている [3]。

本稿では、ユーザの未知の書籍に対する興味を惹きつけ、ユーザが書籍の推薦理由について納得感を得られるようにするために、推薦フレーズとなる文をレビュー中から抽出し、書籍と同時に提示する手法を提案する。推薦フレーズをユーザの趣味・嗜好や注目している観点に基づいて抽出することで、ユーザに「刺さる」ようなフレーズとなり、推薦された書籍に対する読書意欲を湧かせることを目指す。また、ユーザに推薦した書籍の推薦理由としてユーザの趣味・嗜好に合った推薦フレーズを提示することができれば、その推薦の説得力が向上しユーザがその推薦を受け入れる可能性も高まると考えられる。

本研究の提案手法の全体像を図 1 に示す。ユーザに「刺さる」フレーズを抽出するために、本手法では、対象レビューに

1 : <https://bookmeter.com/>

2 : <https://booklog.jp/>

3 : <https://www.honzuki.jp/>

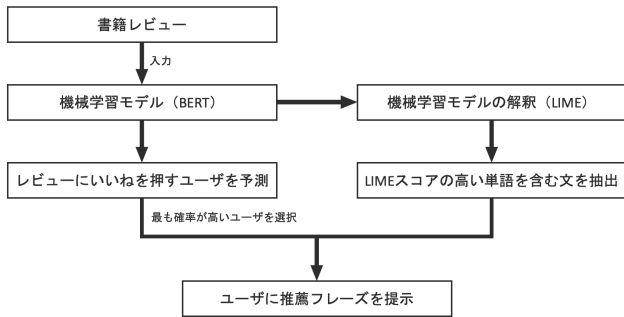


図 1 提案手法の全体図

「いいね」を押すのはどのユーザであるかを機械学習モデルを用いて予測し、ユーザの趣味・嗜好や注目している観点をモデルに学習させる。そして、説明可能な AI の技術の一つである LIME を用いて機械学習モデルを解釈し、レビュー中のどの単語が「いいね」予測の結果に影響をもたらしたのかを予測する。最後に予測された単語を含むレビュー文をレビューから抽出することで、それをユーザに「刺さる」推薦フレーズとして提示する。

本稿の構成は以下の通りである。第 2 章で関連研究について述べる。第 3 章で関連手法である説明可能な AI について述べる。第 4 章で提案手法について述べる。第 5 章で評価実験の結果及び考察を述べる。最後に第 6 章でまとめと今後の課題について述べる。

2 関連研究

近年、アイテム推薦の手法が多く提案されており、またユーザがより推薦に納得できるようなシステムについての研究が増加している。

2.1 ユーザレビューを用いた推薦に関する研究

本節では、ユーザレビューを利用したアイテム推薦に関する研究について説明する。

楠見ら [4] は、ユーザの好みの映画のレビューを分析し、その映画のレビューに含まれる要素を映画の特徴ベクトルとして算出することで、同じ要素を多く含んでいる類似映画の推薦手法を提案している。レビューを用いることで、ユーザの映画ジャンルの好みを推定し、ユーザの嗜好に合わせた映画の推薦を可能にしている。

宮本ら [5] は、単純な協調フィルタリングによる書籍推薦ではなく、ユーザが書籍に対して抱いた印象を考慮した推薦を行うために、レビュー中から印象を表す語を抽出しレビュー間の類似度を得た後、共通レビュアーの投稿したレビュー間でのみ類似性を図ることで、同一の観点に基づく類似書籍を推薦する手法を提案している。しかし、宮本らの手法では、書籍間の類似度がレビューを書いたユーザごとに異なってしまう、同一ユーザがレビューを作成した書籍どうしでなければ類似度は計算することができないという問題点が存在する。本研究では、書籍間の類似度に囚われず、より柔軟にユーザ特徴を得るために、機械学習モデルを用いた学習を行うことでユーザが書籍に対し

抱いた印象を考慮する。

2.2 推薦理由の提示に関する研究

本節では、アイテムを推薦した際の理由の提示、また機械学習モデルの説明可能性に関する研究について説明する。

折原ら [6] は、近年の情報推薦システムは受動的であり自己決定感が少ないことに注目した推薦手法を提案している。この手法では付与された推薦理由の精度が高いとは言えないが、推薦理由が付与されている場合の方が推薦システムに対しての満足度が上昇する傾向があることを示している。

小野川ら [7] は、機械学習の解釈手法である LIME 及び SP-LIME を用いて、商品・サービスのレビューから競合商品・サービスと比較された特徴を抽出する手法を提案している。しかしこの手法には、3 つ以上の商品・サービスについてそれぞれの特徴を抽出できないという問題点がある。

森澤ら [8] は、推薦理由の説明文を自動生成するために、Factorization Machines を用いたユーザ属性に基づく推薦を行った後、LIME の推薦モデルへの適用による特徴量の抽出によって推薦理由文の生成を行う手法を提案している。しかし、この手法ではユーザに提示する推薦理由として性別や年齢といったユーザ属性を用いており、ユーザ属性が同一のユーザの嗜好が同一になることを仮定している。それに対して本研究では、ユーザ属性を利用せず、ユーザが行った「いいね」に基づいてユーザに適した推薦理由の提示を行う。これにより、ユーザがより推薦結果に納得でき、ユーザに「刺さる」推薦理由の提示が可能になると考える。

2.3 キャッチコピー生成に関する研究

本節では、アイテムを推薦する際のキャッチコピーの自動生成に関する研究について説明する。キャッチコピーは単にユーザの購買意欲を煽るだけのものではなく、パーソナライズされた推薦理由の説明を含めることでより有効な推薦につなげることも可能になる。

山根ら [9] は、キャッチコピーの特徴の分析を行い、SNS 上におけるユーザの嗜好を反映させたキャッチコピーの自動生成手法を提案している。福田 [10] は、Autoencoder を用いて、ランダムな文生成ではなく広告対象商品のキーワードによって条件づけられた広告文の生成を行う手法を提案している。

一方で、本研究ではユーザの注目単語を抽出する手法を提案している。そのため、仮に条件キーワードとして本研究の提案手法によって抽出された単語を入力できれば、広告文としての機能を持ちつつユーザの納得感を得られるような文生成への応用も考えられる。

3 説明可能な AI

3.1 説明可能な AI の概要

近年、機械学習技術の発達により、日常生活の様々な場面で AI が積極的に活用されるようになった。それに伴い機械学習モデルは巨大化、複雑化している。その結果、様々な予測問題で精度の向上が図られたものの、機械学習モデルがなぜそのよ

うな予測結果を出力したのかについて、その理由を人間が理解することが困難になった。この問題を解決するために、機械学習モデルに対して「説明」を行うためのアルゴリズムを用いたシステムである「説明可能な AI (XAI: Explainable AI)」の研究が、アメリカの DARPA⁴をはじめとした様々な組織で取り組まれている。

こうした背景のもと、「なぜこのアイテムが推薦されたのか」という推薦の理由を示すことを目的とした説明可能な推薦システムに関する研究も活発に行われている [3]。本研究では数多くある説明可能な AI の技術のうち、様々な機械学習モデルに適用可能であり、代表的な手法の一つである LIME を用いて機械学習モデルの説明を試みる。

3.2 LIME

LIME (Local Interpretable Model-agnostic Explanation) は Ribeiro ら [11] によって提案された機械学習の解釈手法である。この手法では、任意の説明対象データに対する機械学習モデルの予測結果に関して、予測に寄与した特徴を抽出することができる。また LIME は説明可能な AI の技術の中でも、説明対象の AI のモデルの種類に制約がないという利点もある。本研究では LIME を用いることで機械学習モデルの解釈を行う。以下、LIME のアルゴリズムについて説明する。

説明対象モデル $f: \mathbb{R}^d \rightarrow \mathbb{R}$ と入力ベクトル $x \in \mathbb{R}^d$ を考える。このとき、解釈可能なモデル群 G から局所的な近似モデル g を利用して、入力ベクトル x に対する説明 $\xi(x)$ を得る。この関係を式で表すと以下ようになる。

$$\xi(x) = \operatorname{argmin}_{g \in G} (\mathcal{L}(f, g, \pi_x) + \Omega(g)) \quad (1)$$

$$\mathcal{L}(f, g, \pi_x) = \sum_{z, z' \in Z} \pi_x(z) (f(z) - g(z'))^2 \quad (2)$$

ここで \mathcal{L} は f と近似モデル g の乖離度を測る指標（損失関数）、 Ω は説明に使用する近似モデル g の複雑度である。まず、予測対象データ x の周辺に近傍データ $z' \in Z$ を生成する。 $z \in \mathbb{R}^d$ は $f(z)$ を得るために z' から変換したものである。また、 $\pi_x(z)$ は生成したデータ z と説明対象データ x の距離を表す関数であり、 $f(z)$ と $g(z')$ との誤差に重み付けを行うものである。損失関数 $\mathcal{L}(f, g, \pi_x)$ と $\Omega(g)$ の和を最小化することによって、最終的に局所的な線形回帰モデル g を説明結果 $\xi(x)$ として出力することができる。

4 提案手法

本論文で提案する、機械学習モデルと LIME を用いた推薦フレーズの抽出手法について説明する。

4.1 使用するデータ

本研究では、ユーザが「いいね」を押したレビューをデータセットとして用いる。関連研究としてユーザの属性（性別、年齢等）によってアイテムの推薦及び推薦理由の提示を行った

例 [8] もあるが、よりユーザに「刺さる」推薦フレーズを提示するためには、年齢や性別のような客観的なユーザ属性だけでなく、具体的な書籍の内容や読者の感想を取り入れることが重要である。そのため、ユーザの趣味・嗜好や注目している観点を直接的に予測するために、「いいね」を押したレビューを教師データとした機械学習を利用する。

4.2 機械学習によるレビューの「いいね」予測

本手法では、ユーザが未知の書籍のレビュー文を入力として、対象とするユーザ集合の中のどのユーザが「いいね」を押す確率が高いのかを機械学習モデルを用いて予測する。図 2 は機械学習モデルによる「いいね」予測の概念図である。機械学習を行う際の教師データは次のように定義する。

まず、レビューの集合 $R = \{r_1, r_2, \dots, r_{|R|}\}$ 、センテンスの集合 $S = \{s_1, s_2, \dots, s_{|S|}\}$ 、ユーザの集合 $U = \{u_1, u_2, \dots, u_{|U|}\}$ を考える。ここでセンテンスとは、レビューを一文ずつに分割したもの（レビュー文）を指す。このとき、ユーザ u が「いいね」を押したレビューの集合を $like(u) \subseteq R$ と表し、レビュー r に含まれるセンテンスの集合を $sen(r) \subseteq S$ と表すものとする。

いま、センテンスが与えられたときに、そのセンテンスに対して「いいね」を押すユーザを予測するという問題を考える。これは、そのセンテンスを含むレビューに対して「いいね」を押したユーザを正解ラベルとし、機械学習モデルを学習させることで実現できる。センテンス s が与えられたときに、そのレビューに対して「いいね」を押したのがユーザ u である確率は、以下の式 (3) で示す条件付き確率として表現できる。

$$P_{like}(u|s) \quad (3)$$

ここで、任意の s に対して

$$\sum_{u \in U} P_{like}(u|s) = 1 \quad (4)$$

が成り立つものとする。

いま、入力 x に対する正解ラベルが y であるとき、教師データを (x, y) と表すものとする。本手法では入力となるセンテンス s に対する正解ラベルは、以下の条件を満足するユーザ u である。

$$(s \in sen(r)) \wedge (r \in like(u)) \quad (5)$$

したがって、教師データ集合 T は式 (6) ように定義できる。

$$T = \{(s, u) | (s \in S) \wedge (u \in U) \wedge (r \in R) \wedge (s \in sen(r)) \wedge (r \in like(u))\} \quad (6)$$

以上のように定義された教師データを用いて機械学習による「いいね」予測を行う。本研究では、比較のために LSTM, CNN, BERT の 3 つの機械学習モデルを用いて学習を行い、それらの精度を検証する。以下、各モデルについて説明する。

LSTM(Long Short Term Memory) は、Hochreiter ら [12] によって提案された機械学習モデルである。本研究で「いいね」予測を行う LSTM モデルは次のような構造である。単語ごとに分割されたレビュー文を入力とし、それを分散表現に変換し

4 : <https://www.darpa.mil/program/explainable-artificial-intelligence>

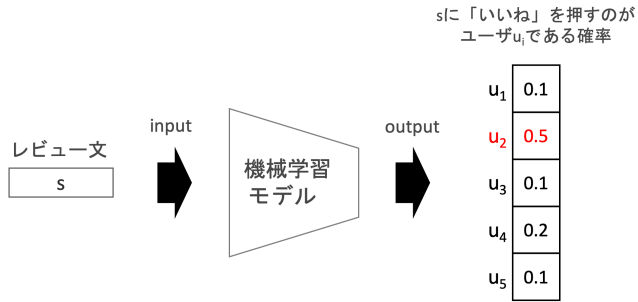


図2 機械学習による「いいね」予測の概念図. 赤字は「いいね」を押す確率が最も高いと予測されたユーザー

たものを LSTM に入力する. その後, 全結合層にソフトマックス関数を組み合わせることで, 入力したレビュー文に対してどのユーザーが「いいね」を押すかの確率を出力する.

CNN(Convolutional Neural Network) は, 主に画像分野の機械学習に用いられることが多いモデルである. 自然言語分野においても, Kim [13] をはじめとした研究で利用されており, 一定の性能を示している. 本研究では Kim の研究を参考に, 「いいね」予測を行う CNN モデルを構築する. まず, 単語ごとに分割された文を入力し, それを分散表現に変換する. すなわち入力した文は, (文の系列長 \times 分散処理の次元数) の大きさのベクトルで表されることになる. そのベクトル表現に対して, 一次元の畳み込み層とプーリング層を適用することで文章の特徴を得る. その後, 全結合層にソフトマックス関数を適用することで, 入力したレビュー文に対してどのユーザーが「いいね」を押すかの確率を出力する.

BERT(Bidirectional Encoder Representations from Transformers) は, Devlin ら [14] によって提案された機械学習モデルである. BERT による多クラス分類問題では, 最終層の Transformer Encoder の [CLS] トークンの出力を, 全結合層へ入力することでクラスラベルを求めることができる. 本研究で「いいね」予測を行う BERT モデルは, 図3のような構造である. 単語ごとに分割された文を BERT モデルに入力し, シーケンスの先頭に付与した [CLS] の出力を得る. この [CLS] トークンの最終的な隠れ層は入力文全体の情報をエンコードしているため, それに全結合層とソフトマックス関数を適用することで, 入力したレビュー文に対してどのユーザーが「いいね」を押すかの確率を出力できる.

上記の LSTM, CNN, BERT の3モデルを用いて精度の比較を行い, 最も精度の高かったモデルに対して次節の LIME の適用を行う. LSTM, CNN の入力時の単語の Embedding には fastText [15] を用いる. また BERT モデルについては, 事前学習済み BERT モデルを fine-tuning することで学習・予測を行う. 事前学習済み BERT モデルには, 東北大学乾研究室による日本語 Wikipedia 学習済みモデル [16] を用いる. 教師データを用いてモデルを学習させた後, ユーザーが未知の書籍のレビューのレビュー文をモデルに入力し, 「いいね」を押すユーザーの確率を求める.

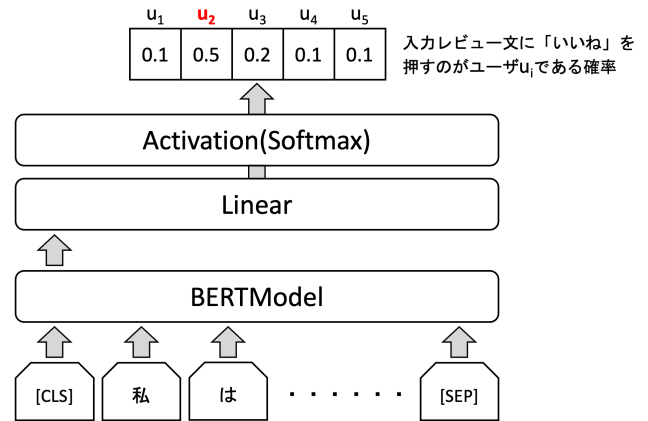


図3 BERT による「いいね」予測モデル

4.3 LIME によるユーザー注目単語及び推薦フレーズ抽出

4.2 節の「いいね」予測において, 任意のレビュー文に対し最も「いいね」を押す確率が高いユーザーを対象に, LIME を用いて予測結果に影響を与えたと考えられる単語をレビュー文から抽出する.

本手法では, 対象とするユーザーに対して, 推薦する書籍があらかじめ決まっているときに, その書籍に関するレビュー文の中からユーザーに「刺さる」フレーズを抽出することを目的とする. そこで, 対象とする書籍のレビューのレビュー文に対して, そのレビュー文に「いいね」を押す確率が最も高いユーザーを予測する. そして, その予測結果に対して LIME を適用することで, 「いいね」ユーザー予測をする際の各単語の影響度合いを計算する. 図4に「いいね」ユーザー予測の結果に影響を与えたとされる単語を色付けし, その影響度合いを可視化した例を示す.

いま, 書籍 b を対象に, ユーザー u に対する単語の影響度合いを予測することを考える. 書籍 b に関するレビュー集合を $rev(b) \subseteq R$ とする. このとき, 対象となるレビュー文集合 $Sen(b) \subseteq S$ を以下のように定義する.

$$Sen(b) = \{s | (s \in sen(r)) \wedge (r \in rev(b))\} \quad (7)$$

また, レビュー文 s に「いいね」を押すユーザー u を予測するモデルに LIME を適用して得られる単語 w の重みを $LIME(w, s, u)$ と表記する. このとき, その単語を含むレビュー文 s に「いいね」を押すユーザーが u であると予測された確率 $P_{\text{like}}(u|s)$ との積を求めることで, レビュー文 s における単語 w に関するユーザー u の単語注目度 $significance(u, w, s)$ を以下のように定義する.

$$significance(u, w, s) = LIME(w, s, u)P_{\text{like}}(u|s) \quad (8)$$

図4を例にすると, ユーザー1の「犯罪」の単語注目度は, LIME のスコア 0.11 と該当レビューに「いいね」を押したユーザーが1である確率 0.47 の積, 0.0517 となる. このようにして, ある1冊の書籍の全レビュー文集合 $Sen(b)$ に対して LIME を適用して, 対象ユーザーに関するユーザー単語注目度を算出する.

次に, レビュー文 s に含まれる単語 w の集合を $W(s) = \{w_1, w_2, \dots, w_{|W|}\}$ とする. このときユーザー u のレビュー文 s

Text with highlighted words

犯罪者を庇ったために人生の歯車が狂い、いろんな人を巻き込み被害者家族だけでなく、加害者家族も苦しめる結果に。

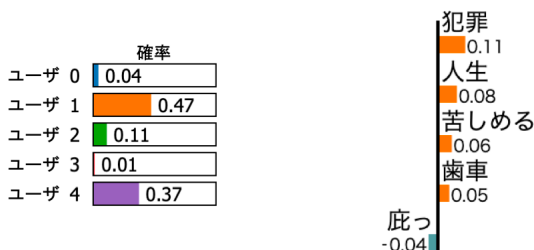


図 4 LIME の実行例. 上から入力したレビュー文, 「いいね」予測の確率 (左下), LIME のスコア (右下)

のスコア $S_{\text{sen}}(u, s)$ を, その文に含まれる単語の単語注目度の最大値として以下のように定義する.

$$S_{\text{sen}}(u, s) = \max_{w \in W(s)} \text{significance}(u, w, s) \quad (9)$$

現在対象としている書籍のレビュー文の S_{sen} を全て算出し, S_{sen} の最も高かった上位 N 件のレビュー文を推薦フレーズとして定義する.

例えば図 5 のように, 「いいね」を押すのがユーザー u であると予測された複数のレビュー文があったとき, それらのレビュー文の単語全てに対して単語注目度を算出することができる. その後, 単語注目度の高かった上位 3 件の単語, 「ビジネス」, 「数学」, 「感動」を含んでいるレビュー文を, ユーザに提示する推薦フレーズとして抽出する.

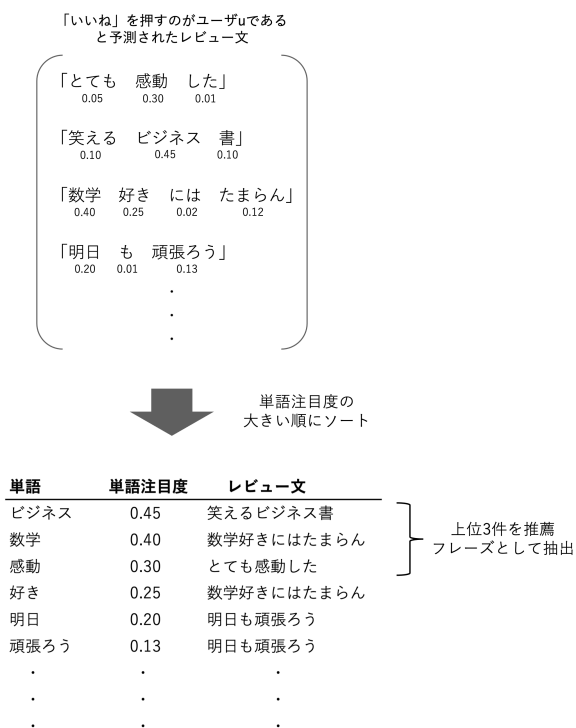


図 5 単語注目度に基づく推薦フレーズ抽出の流れ ($N = 3$ のとき)

5 評価実験

5.1 機械学習を用いた「いいね」予測実験

本節では機械学習を用いたレビューの「いいね」予測の結果を述べる. データセットとして, 書籍レビューサイトの「ブックログ」に投稿されたレビューを使用した. 対象とするユーザー数は 5 人, 各ユーザーが「いいね」を押したレビューの合計は 757 件であった. また「いいね」を押したレビューを文ごとに分割することで, 学習データ数を 4457 件とした. 学習用データのうち 80 % を訓練用データ, 10 % を検証用データ, 10 % をテストデータに分割した. 入力するトークンの最大長は 256 とした.

レビュー文を機械学習モデルに入力するために前処理を行った. レビュー文の分かち書きには MeCab [17] を Python から使用する際のラッパーライブラリである fugashi を用いた. この際, 辞書として ipadic を用いた. レビュー中の URL の記述は削除し, 英語・記号については全て半角文字に統一した. また数字については全て 0 に変換し, 数字による影響が出ないようにした.

次に, それぞれのレビュー文を入力として, 対象ユーザー 5 人それぞれがレビュー文に「いいね」を押したユーザーである確率を出力する学習・予測を行った. 機械学習モデルのパラメータとして, LSTM と CNN の単語の分散表現は 300 次元, バッチサイズは 128 とし, CNN のフィルタ数は 250, カーネルサイズは 3, 1 次元畳み込み層の活性化関数には relu を使用した. 「いいね」予測を行う際には, 各モデルで検証データの損失が最も低かった際のモデルを用いた.

各モデルの結果は表 1 の通りである. LSTM では 4 エポック学習し, 45.44 % の accuracy を得た. CNN では 7 エポック学習し, 49.48 % の accuracy を得た. BERT では 3 エポック学習し, 51.46 % の accuracy を得た. この実験では BERT モデルの精度が最も高かったため, 次節の LIME の適用では BERT モデルを対象とした.

表 1 機械学習モデルごとの accuracy

Model	accuracy
LSTM	0.4544
CNN	0.4948
BERT	0.5146

5.2 LIME による推薦フレーズ抽出の評価実験

5.2.1 実験方法

LIME によって抽出された単語を含む文を推薦フレーズとしてユーザーに提示し, その有効性を確かめるための実験を行った. 被験者の人数は 5.1 節にて使用したデータを持つ 5 人のユーザーに参加してもらった. 被験者にとって未知の書籍 10 冊を推薦書籍としてあらかじめ定め, それぞれの書籍に対して提案手法による推薦フレーズの抽出を行った. 被験者に提示する推薦フレーズの本数は, 各書籍で単語注目度が高かった上位 3 件, 計 30 フレーズとした. 被験者には推薦フレーズの他に, 書籍のタイ

トル、著者、表紙の画像も同時に提示した。また、比較手法として次節の手法で抽出された推薦フレーズも被験者に提示した。

5.2.2 比較手法

提案手法の有効性を確かめるために RP-others, RP-tfidf の2つの比較手法を用いた。

RP-others では、提案手法によって抽出された推薦フレーズが、対象となるユーザごとにパーソナライズされているのかを評価することを目的とする。注目している対象ユーザ以外の4人のユーザに対して提示された推薦フレーズのうち、1冊につき3フレーズ、計30フレーズをランダムに抽出して対象ユーザに提示した。提案手法と同様に、提示された推薦フレーズを評価してもらった。

RP-tfidf では、tfidf 値を利用してレビュー文のスコア付けを行った。tfidf 法とは文書に含まれる単語の重要度を評価する手法の一つであり、tf 値 (Term Frequency, 単語の出現頻度) と idf 値 (Inverse Document Frequency, 単語の逆文書頻度) を掛け合わせることで求められる。4.2 節の手法によって、入力した全てのレビュー文を「いいね」を押す確率が最も高いユーザごとに分類した後、そのユーザごとに分類されたレビュー文の集合を1つの文書とみなすことで、単語の tfidf 値を算出した。その後、レビュー文ごとに単語の tfidf 値の平均を算出し、平均値の高かった上位3件のレビュー文をユーザに提示する推薦フレーズとした。以下は具体的な計算方法である。

ユーザ u が「いいね」を押すと予測されたレビュー文 s に含まれる単語 w の集合を $W(s) = \{w_1, w_2, \dots, w_{|W|}\}$ とする。このときレビュー文 s のスコア $S'_{\text{sen}}(s)$ を、その文に含まれる単語 w の tfidf 値 $tfidf(w)$ の平均値として以下のように定義する。

$$S'_{\text{sen}}(s) = \frac{\sum_{w \in W(s)} tfidf(w)}{|W(s)|} \quad (10)$$

全てのレビュー文に対して S'_{sen} を算出し、 S'_{sen} の最も高かった上位 N 件のレビュー文を RP-tfidf における推薦フレーズとした。本実験では、提案手法と比較手法で同じレビュー文数になるように、上位3件までを推薦フレーズとした。

5.2.3 評価項目

提案手法, RP-others, RP-tfidf の3つの手法で抽出された各30件、合計90件の推薦フレーズを被験者にランダムな順序で提示し、評価を行った。

評価の際の質問項目は以下の4つである。回答は5段階のリッカート尺度 (5.非常にそう思う, 4.ややそう思う, 3.どちらとも言えない, 2.あまりそう思わない, 1.全くそう思わない) を用いた。

- 質問1 : この推薦フレーズはあなたに「刺さる」ものだった
- 質問2 : この推薦フレーズによってこの書籍に興味を持った
- 質問3 : この推薦フレーズは分かりやすかった
- 質問4 : この推薦フレーズは役に立つ

その後、提案手法と比較手法で質問の回答に有意差がみられるかの検定を行った。有意差を求める際は、Welch の t 検定及び Mann-Whitney の U 検定を用いて、それぞれで p 値を算出した。

5.2.4 結果

表2は『博士の愛した数式』⁵を推薦する書籍としたときの各ユーザの単語抽出結果、単語注目度 *significance* 及び推薦フレーズ抽出の結果の例である。図6は各質問に対する各評価値の回答の割合、図7は各質問に対する評価値の平均値及び t 検定による有意差の有無の結果を示したグラフである。

また、提案手法と各比較手法の有意差の有無を調べるために行った Welch の t 検定及び Mann-Whitney の U 検定の結果を表3, 表4に示す。表より有意差の有無の結果は、両検定手法とも全て同じ結果になった。質問ごとに p 値を求めると、RP-others については、質問1では両検定とも有意水準5%で有意であることが確かめられた。一方、質問2~4では両検定とも p 値が0.05以上となり、提案手法と RP-others で有意差は認められなかった。RP-tfidf については、全ての質問で p 値が0.05以下となり、両検定において有意水準5%で有意であることが確かめられた。特に質問3に関しては有意水準1%で有意であった。

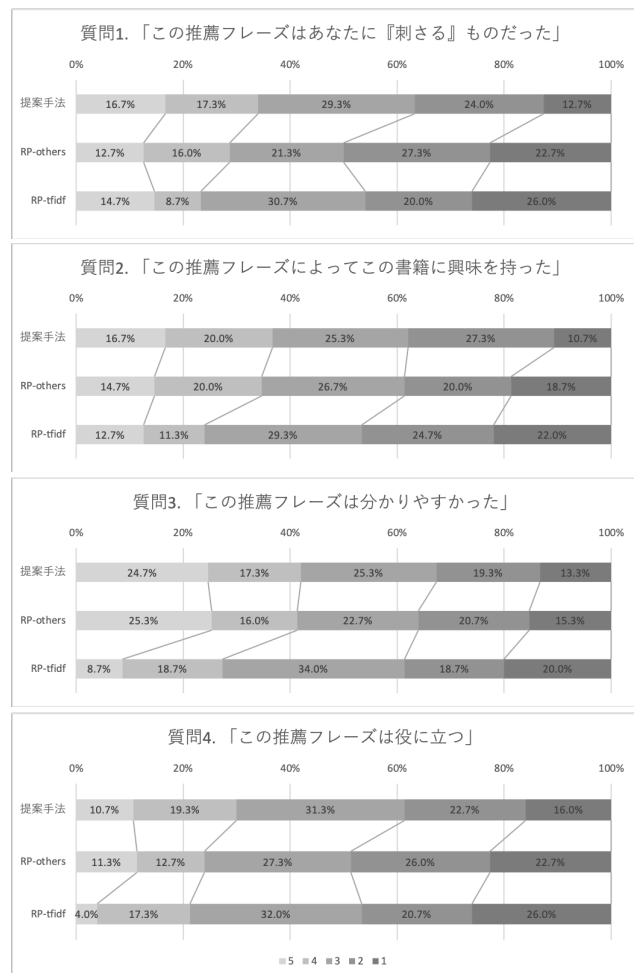


図6 各質問の回答の割合 (5.非常にそう思う, 4.ややそう思う, 3.どちらとも言えない, 2.あまりそう思わない, 1.全くそう思わない)

5: 『博士の愛した数式』 小川洋子著. 新潮文庫. 2005

表 2 提案手法による『博士の愛した数式』の結果の例

ユーザ	抽出単語	単語注目度	推薦フレーズ
1	恋	0.47234	帰ってきたら恋が終わってしまう気がするから。
	愛し	0.43323	記憶を失っても愛し続ける。
	温まる	0.41626	ころころ温まる物語。
2	数学	0.44352	数学の美しさやおもしろさをこんなに情感豊かに表現する本はないかもしれません。
	数学	0.42791	完璧になんでもできる人ではないけれど、数学と幼い者への愛情あふれる博士の物語。
	数学	0.42756	数学好きな私としてはとても面白く、かといって数学数学してるかというところでもない。
3	行動	0.21494	人を思いやり、純粋な気持ちで起こす行動は、なぜこんなに心を打つのだろう。
	近づける	0.19530	そうすることで、自分も素敵な人間に近づけると思う。
	入力	0.18578	ISBN ナンバを入力している時、何か不思議な気分にとらわれました。
4	タイガース	0.08552	思っていたよりずっと明るく楽しいおしゃべりで、大ファンであるタイガースのこと、自作のことをイキイキと語られていて、聞いているこちらまで気持ちが弾むよう。
	福	0.08142	福ちゃん。
	江夏	0.07211	江夏は偉大だ。
5	博士	0.35027	博士が自分はちっぽけな存在などと思っているところがとても切なくなる。
	思っ	0.32729	そう思っていた。
	博士	0.29224	博士にも普通の人と同じように振舞いたいと言うプライドが見え隠れしているがそれはものすごいストレスを博士に強いていたに違いない。

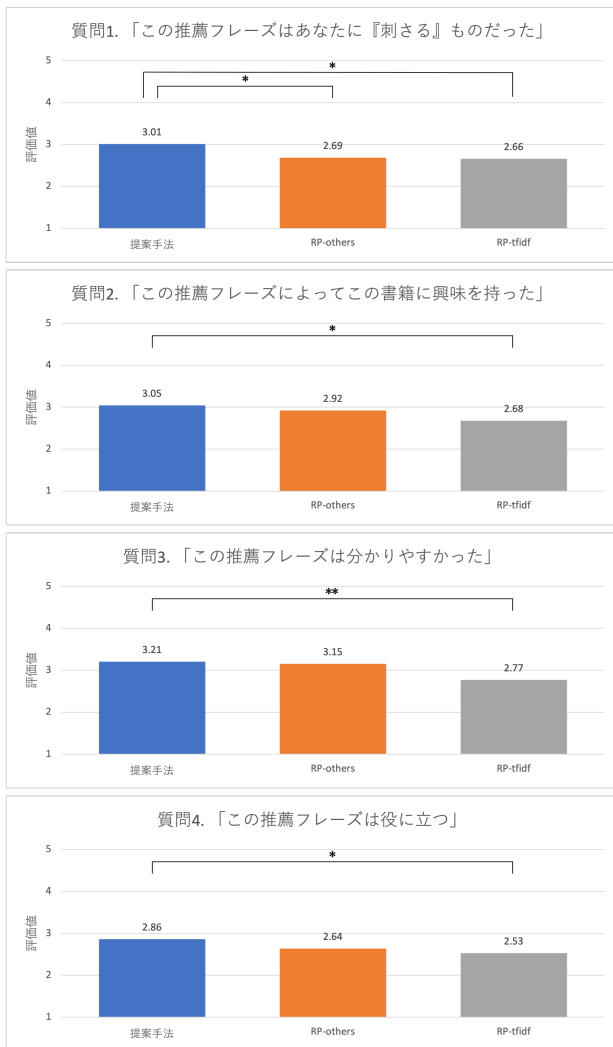


図 7 各質問の評価の平均値 (* p<0.05, ** p<0.01)

5.3 考察

評価実験の結果について、図 6 より、提案手法のほうが比較

表 3 t 検定による p 値 (提案手法との有意差)

	質問 1	質問 2	質問 3	質問 4
RP-others	0.030	0.395	0.739	0.127
RP-tfidf	0.020	0.013	0.004	0.016

表 4 U 検定による p 値 (提案手法との有意差)

	質問 1	質問 2	質問 3	質問 4
RP-others	0.028	0.451	0.748	0.105
RP-tfidf	0.020	0.016	0.008	0.029

手法よりも各質問に対する評価が高い傾向がある。

RP-others について、有意差検定の結果より、質問 1 では提案手法の回答に対して有意差があったが、質問 2~4 では有意差があるとは言えないことが示された。しかしながら RP-others を実施した目的が、抽出された推薦フレーズがユーザごとにパーソナライズされているかを調べるためのものであることを考慮すると、質問 2~4 に有意差がなくても、質問 1 (推薦フレーズのパーソナライズ) では提案手法との有意差があることから問題ないと考えられる。よって、本研究の提案手法は、RP-others と比べて有効であったといえる。

また RP-tfidf について、有意差検定の結果より、全ての質問で提案手法の回答に対して有意差があることが示された。したがって、提案手法によって抽出された推薦フレーズは、RP-tfidf よりも推薦フレーズのパーソナライズ化 (質問 1)、推薦の効果 (質問 2, 4)、フレーズの理解度 (質問 3) の全ての観点において効果が高かったと考えられる。

ただ、ほとんどの質問で回答の割合が最も高かった評価値は 3 (どちらとも言えない) であり、評価の平均値も 3 前後であった。また、評価値 3, 2, 1 (どちらとも言えない、あまりそう思わない、全くそう思わない) と回答した割合も全体のおよそ 60% であることから、本研究の提案手法は比較手法に対しては有効であったと言えるが、提案手法の結果のみを考えた場合に

表 5 その他の抽出された推薦フレーズの例

1	だけど、読者にとっても彼女たちの話はかなり衝撃的。
2	それが全ての根源的問題である。
3	黒澤のキャラがやっぱり好き。
4	侮るなかれ。

は、適切な推薦フレーズを多く提示することができていたとは言えない。

次に、被験者ごとの注目単語及び推薦フレーズの抽出結果の例を具体的に見る(表 2)。ユーザごとに抽出された単語が異なっており、どのような単語に注目しているのかの傾向を推測できる。例えば、ユーザ 2 は「数学」という言葉に、またユーザ 4 は野球関連の言葉に強く注目していることがわかる。特にユーザ 2 についてはその他の書籍での単語抽出結果においても、「論理」「収束」「物理」など自然科学関連の単語が多く抽出されており、ユーザの注目している観点の傾向を把握できる。

また、被験者の中には「一言目が接続語、指示語で始まっている推薦フレーズ(例:表 5-1, 5-2)や、固有名詞を含んだ推薦フレーズ(例:表 5-3)は分かりづらい」という意見があった。これは、推薦フレーズがレビュー中からそのまま抽出された文であることが原因として考えられ、複数の文からなるレビューでは文章の前後の文脈がないと意味が伝わらなかつたり、書籍を読んでいないと分からない固有名詞が出てきたりしたことに繋がると推測される。このような問題は提案手法、比較手法のどちらにも発生していた。

一方で「一言だけの短い推薦フレーズ(例:表 5-4)は、意味がよく分からなくても深読みして、逆に興味が出てしまった」という意見もあった。そのため、一般的に意味を理解しやすいフレーズだけが推薦フレーズとして適切であるとは言い難く、抽象度の高いフレーズを好むユーザもいることが分かった。ゆえに推薦フレーズの長さ及びフレーズの理解のしやすさ度合いも含めてユーザごとにパーソナライズすることができれば、さらにユーザに「刺さる」推薦フレーズを提示することが可能になると考えられる。

6 おわりに

本研究では、ユーザにアイテムを推薦する際に推薦フレーズを提示する手法として、機械学習によるユーザの「いいね」予測によってユーザの趣味・嗜好や注目している観点を捉えた機械学習モデルを構築した後、機械学習モデルを解釈するアルゴリズムである LIME を適用することで、ユーザの注目単語を含む推薦フレーズを抽出する手法を提案した。提案手法により、ユーザごとの特徴を捉えた推薦フレーズを提示することができ、tfidf 値による単語の重みづけを利用した手法と比べて、よりユーザに「刺さる」推薦フレーズを抽出することができた。ただし、抽出された推薦フレーズの中には、意味が分かりづらく、ユーザに刺さらなかったものも一定数含まれていることも示された。

今後の課題として、よりユーザにわかりやすく、強い印象を

与えられるような推薦フレーズを提示するために、レビューから文を直接抽出するのではなく、推薦フレーズを自動生成できるようなシステムの開発を検討している。それにより、柔軟な文生成が可能となり、更なる宣伝効果を期待できるような推薦フレーズをユーザに提示することが期待できる。また、本研究では推薦する書籍をあらかじめ指定していたが、今後は既存の推薦システムと本研究の手法を組み合わせることで、より満足度の高い推薦を行うシステムの開発にも取り組みたい。

謝 辞

本研究は JSPS 科研費 19H04219 の助成を受けたものです。

文 献

- [1] Michal Kosinski, David Stillwell, and Thore Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 2013.
- [2] Alexandra Vultureanu-Albiși and Costin Bădică. Recommender Systems: An Explainable AI Perspectives. Computer and Electronics University of Craiova, 2021.
- [3] Yongfeng Zhang and Xu Chen. Explainable Recommendation: A Survey and New Perspectives. In *Foundations and Trends in Information Retrieval*, 2020.
- [4] 榎見圭司, 北山大輔. ユーザレビューを用いた全体的・部分的観点の類似に基づく映画推薦. DEIM2017 B5-4.
- [5] 宮本達矢, 北山大輔. 共通レビューの観点の類似性に基づく書籍推薦手法. DEIM2018 F2-5.
- [6] 折原レオナルド賢, 橋山智訓, 田野俊一. 推薦理由を提示する情報推薦システム. The 29th Annual Conference of the Japanese Society for Artificial Intelligence, 2015.
- [7] 小野川稔之, 折原良平, 清雄一, 田原康之, 大須賀昭彦. 機械学習モデルの解釈手法による競合サービスと比較したレビュー分析. 日本ソフトウェア科学会第 36 回大会, 2019.
- [8] 森澤竣, 山名早人. 機械学習モデルの解釈手法を用いたアイテム推薦理由の説明文の生成. DEIM2020 E6-2.
- [9] 山根宏彰, 萩原将文. ウェブにおけるユーザの嗜好を反映させたキャッチコピー自動生成システム. 日本感性工学会論文誌 Vol.13 No.4 (特集号), pp. 493-500. 日本感性工学会, 2014.
- [10] 福田宏幸. キーワード条件つき変分 autoencoder による広告文生成. The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019.
- [11] Marco Tulio Ribeiro, Sammer Singh, and Carlos Guestrin. "Why Should I Trust You?" Explaining the Predictions of Any Classifier. 2016.
- [12] Sepp Hochreiter and Jürgen Schmidhuber. Long Short-term Memory. In *Neural Competition, Volume 9, Issue 8*, pp. 1735-1780, 1997.
- [13] Yoon Kim. Convolutional Neural Networks for Sentence Classification. New York University, 2014.
- [14] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Google AI Language, 2018.
- [15] fastText. <https://github.com/facebookresearch/fastText>.
- [16] 訓練済み日本語 BERT モデル. 東北大学乾研究室. <https://github.com/cl-tohoku/bert-japanese>.
- [17] MeCab. <https://taku910.github.io/mecab/>.