著者に注目した映画レビュー推薦に適した言語モデルの生成

吉田　修平† 　荘司　慶行† 　藤田　澄男†† 　Martin J. Dürst†

† 青山学院大学理工学部情報テクノロジー学科　〒 252–5258 神奈川県相模原市中央区淵野辺 5–10–1
†† ヤフー株式会社　〒 102–8282 東京都千代田区紀尾井町 1-3 東京ガーデンテラス紀尾井町 紀尾井タワー
E-mail: †shuhei@sw.it.aoyama.ac.jp, ††{shoji,duerst}@it.aoyama.ac.jp, †††sufujita@yahoo-corp.jp

あらまし　本論文では、レビューの著者に注目してファインチューニングした BERT を用いることで、ある個人の価値観に合いそうなレビューを推薦する手法を提案する。映画レビューサイトでは、様々な人々が，それぞれ異なった価値観からレビューを書いて投稿しているために、著しく価値観の異なるレビューは映画を見るかどうかの判断に使いづらい。本研究では、レビューに反映される価値観は、著者に固有の言及している観点や文体などに起因すると考えた。そこで、ある 2 つの任意の映画に対するレビュー文を与えた際に、それらのレビューの著者が同一であるかを推定するタスクで BERT モデルをファインチューニングした。こうして作成された言語モデルは、著者の価値観を反映した観点や文体を識別するようにレビューをベクトル化すると考えられる。2 件のレビュー文がそれぞれどう似ているかをラベル付けするクラウドソーシングによる実験から，提案する言語モデルで生成されたベクトルが確かに個人の文体を強く反映していることを確認した.

キーワード　情報推薦，レビュー，BERT, Self-Supervised Learning

1 はじめに

近年，映画のサブスクリプションサービスやレビューサイトの普及により，インターネット上に投稿されるレビューの総数も，それらを目にする機会も増加してきている．しかしながら，人にとって年齢や性別，その他の様々な属性から共有される文脈や価値観が異なるため，参考になるレビューは，人によって異なる場合が多い．例えば，ある映画に関して「出演している俳優を目当てに鑑賞した人」にとって「脚本を重視する映画マニア」のレビューが参考になるとは限らない．しかし多くの映画レビューサイトにおいては，ユーザにパーソナライズされた映画推薦は行われているものの，そのレビュー自体は単純な投稿日時や，評価の順序で提示されており，パーソナライズされたものになっていないのが普通である．

また映画レビューサイトのニーズとして，その映画を見るかどうかの判断のためという目的に加えて，自分が既に見た映画について他のユーザの意見が気になるという答え合わせ的な目的としても使われることが多い．そこで本研究では，答え合わせとしてのレビュー観覧のニーズに合わせて，あるユーザによって書かれたレビュー文を元に，そのレビューが対象としている映画のレビューを推薦する手法を提案する．

レビュー自体を対象とする推薦はアイテムの推薦とは異なり，それ自体が点数化されることが少ないために直接的な教師あり学習の枠組みでは扱うことができないという技術的な困難性が存在する．そのため間接的なファインチューニング用のタスクを設定する必要がある．まず，そのための前提として，こういった文脈や価値観が，レビューの文体や用いられている語彙に反映されていると考えた．そして Sentence-BERT を，ある 2 つの任意の映画に対するレビュー文を与えた際に，それらのレビューの著者が同一であるか否かを推定するというタスクによってファインチューニングを行った．

こうして作成された言語モデルは，著者の価値観を考慮してレビューをベクトル化すると考えられる．そのモデルによってベクトル化されたレビュー文を元となるレビューとの類似度が高い順序で提示することで映画レビューの推薦を行う．学習には Yahoo!映画に投稿されているレビュー文を用いる．手法の評価としてはレビュー推薦では，Yahoo!映画の実データを用い，クラウドソーシングによる大規模な評価実験を行った．被験者は，1 件のレビュー文と，提案手法を含む 4 つの手法で計算された推薦結果上位のレビュー文を読み，それぞれどのような観点で類似しているかをラベル付けを行った．そして，結果からこの言語モデルの特性について議論した．

本論文は全 6 節から構成されている．第 2 では本研究と関連する研究として言語モデルと情報推薦について論じる．第 3 章では，本研究で提案する，同一著者によって書かれたレビューであるかを推定するタスクを通して言語モデルをファインチューニングする方法について述べる．第 4 節では，提案手法によって作成されたレビューのベクトルの性質と，その有用性について評価するための実験および結果について述べる．その結果について第 5 節で考察し，第**??**節で総括する．

2 関連研究

本研究では，個人の嗜好に合わせたレビューを推薦するために，ファインチューニングした BERT モデルを用いる．そこで，目的の類似する研究として分散表現を用いた情報推薦および BERT を用いた情報推薦について紹介し，論じる．加えて，使用した技術として，BERT のファインチューニングについて関連研究について述べ，関連性を示す．

2.1 分散表現を用いたアイテム検索・推薦

本研究は，映画のレビュー文を高速に類似度を比較できるような軽量な形で分散表現化することで，レビュー，アイテムの推薦を行う．このような分散表現を用いた推薦に関する研究は，一般的に行われてきている．

例として，Barkan ら [1] は Word2Vec として知られる分散表現化の手法である Skip-gram with Negative Sampling (SGNS) をアイテムベースの協調フィルタリングに応用する Item2Vec という手法を提案している．この手法では，アプリケーションサイトにおいてアプリの購入ログを，音楽配信サイトにおいて再生ログに SGNS を適用することで，アイテムの分散表現を得ることを可能にしている．Phi ら [2] は EC サイトにおける協調フィルタリングによるアイテム検索に分散表現を用いた手法を提案している．この研究では，アイテムを単語，ユーザのセッションを文章として扱い，アイテムに対しては Word2Vec を，ユーザに関しては Doc2Vec を適用して，実際にそれぞれを分散表現として表している．

古典的な協調フィルタリングによる推薦には，データ量が少ない場合に正確な推薦が行えないことや，コールドスタート問題などの欠点が存在する．これらを解決する手法として Liu ら [3] らは Doc2Vec を用いてテキストから映画を分散表現化することで，その類似度を協調フィルタリングに適用する手法を提案している．実験では推薦結果が精度, 再現率ともに向上することが確認されている．

2.2 BERT による類似度比較

前節に記した多くの手法の基礎となる Word2Vec は単語に対して文脈非依存の分散表現を得るための手法である．本研究では，文脈を考慮した高精度な分散表現を用いて推薦を行うため Devlin ら [4] の提案した BERT を用いた．

BERT における STS は，構造上単純な 2 文書の比較では高い性能が得られる反面，計算量が膨大になる．そのため，意味的な類似性検索や，クラスタリングのような教師なし学習には適していない．また BERT の出力の Average-Pooling による類似度計算では GloVe を使用したものより精度が劣ることが一般に知られている．

こうした問題への対応策の例として，Reimers ら [5] は Siamese network [6] と呼ぶ 2 つのニューラルネットワークを使って文書の分散表現を得るためのモデルを提案している．この手法では，高速な類似度計算が可能な分散表現化を可能にしている．

2.3 BERT による推薦

自然言語処理分野のタスクで高い精度を持つことを示した BERT は，近年では情報検索や情報推薦など様々な応用的な研究に用いられている．BERT においては，事前学習後に知識をパラメータに暗黙的に有していることも示されている．Penha ら [7] は事前学習済みの BERT が本，映画，音楽などの項目についてどの程度知識を有しているか，また推薦における協調ベースの知識よりも内容ベースの知識が多いことなどを示している．

Malkiel ら [8] は BERT によって得られる分散表現を用いて，プロのワインレビューに基づくワイン推薦の手法を提案している．ここで提案される学習の手法は人手によるラベリングを必要としない自己教師あり学習である．

本研究でも，ラベリングを必要としない間接的なファインチューニングタスクによって学習を行ったモデルを用いて推薦を行う手法を提案する．

3 提 案 手 法

本研究では，コサイン類似度の比較に用いることのできる，レビューの分散表現を得るためのモデルと，そのためのファインチューニング方法を提案する．そしてそこから得た分散表現を用いてレビュー自体の推薦を行う手法を提案する．

本手法は本来は言語非依存だが，本論文では日本語で書かれた映画レビューセットを対象として実験を行ったため，本節では日本語を例にとって手法を説明する．

3.1 モ デ ル

学習済み BERT モデルをファインチューニングすることで，推薦に必要となるコサイン類似度で比較可能な分散表現を得るためのモデルを作成する．言語モデルは Sentence-BERT の分類問題に対応したものを用いる．Sentence-BERT では BERT の出力にプーリング操作を追加したもので，論文での実装は Max-Pooling，Average-Pooling があるがここでは Average-Pooling を用いた．生成された文の埋め込みが意味的に意味があり，コサイン類似度と比較できるように重みを更新するために siamese network を用いる．

モデルの概略を図 1 に示す．このモデルでは，入力として 2 つのレビュー文を受け取る．入力されたそれぞれのレビューは BERT で単語単位で分散表現化されるので，Average-Pooling することで 1 つのベクトルとして表せる．こうして生成された 2 つのレビュー文の分散表現どうしを，1 本のベクトルに結合する．この際，ベクトルの末尾に，2 つのレビューの分散表現の各次元の差の絶対値をとったものを，追加で結合する．このようなモデルについて，実際に，同一の著者によるレビューであるかどうかを学習し，ファインチューニングを行う．

3.2 損 失 関 数

図で示される文書の埋め込みである u と v，及び差分ベクトルの各要素の絶対値をとったものを連結し，学習可能な重みをかけたものに softmax 関数を通してその交差エントロピー損失を最適化する．具体的には，損失値 o は，

$$o = softmax(W_t(u, v, |u - v|)) \tag{1}$$

として表される．この際，それぞれの損失にかかる係数 W_t は，n を埋め込み表現の次元，k をラベル数とした際，

$$W_t \in R^{3n \times k} \tag{2}$$

と表せる．

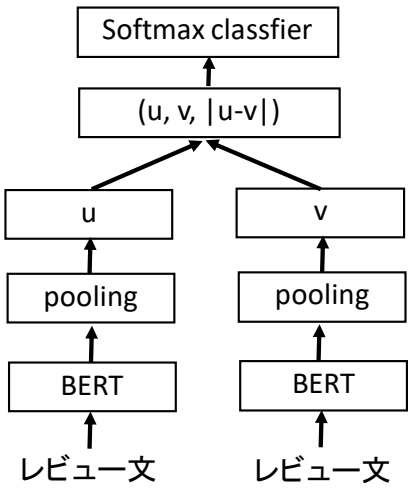

図 1　Sentence BERT を参考に構築した提案手法のモデル．2 つのレビュー文を BERT でベクトル化し，プーリングしたベクトルを結合したうえで，それぞれのレビューが同一の著者に書かれたかどうかを推定するタスクで学習する．

3.3　テキストの前処理

使用するレビュー文はモデルへ入力する前に前処理を行った．括弧の除去，URL の除去，全角空白の除去，単語の正規化を行った．単語の正規化では単語の文字種の統一，つづりや表記揺れの吸収といった単語を置き換える処理を行った．

3.4　ファインチューニング

ファインチューニングとしては正例として入力の 2 つのレビューが同一ユーザによって書かれたものを，負例として入力の 2 つのレビューが異なるユーザによって書かれたものを用いて，それを推定するタスクを実行する．

3.5　レビュー推薦

本研究では，ユーザが既に投稿したある映画のレビューを元に，同タイトルについて書かれたレビューを推薦する手法を提案する．つまり想定されるシチュエーションというのはユーザが既に見た映画について他のユーザはどのような意見をもっているのか知りたいという場合である．

推薦するべきレビューが対象とする映画タイトルについて書かれたレビューの分散表現を提案手法における学習済みモデルによって得る．そして元となる映画レビューの分散表現とのコサイン類似度をすべて計算し，そのスコアが高いものを推薦するべきレビューとする．

4　評 価 実 験

本章では，提案手法の有用性を示すために用いたデータセットと，そのデータセットを用いて行った評価実験の詳細について述べる．提案手法，及び評価対象となる手法によって得られた映画レビューの分散表現から，同じ映画についてのレビューの中から類似度の高い 5 つを抽出し，その特性を明らかにするための 0 つの質問についてクラウドソーシングによる大規模なアンケートを実施した．

表 1　クレンジング後のデータセット内での一人当たりの投稿レビュー数と人数（概算）

投稿数	人数	比率
1〜5 件	307,500	86.5%
5〜10 件	23,500	6.6%
10〜20 件	12,000	3.4%
20〜50 件	7,700	2.2%
50〜100 件	2,600	0.7%
100〜200 件	1,200	0.3%
200〜500 件	600	0.2%
500〜	300	0.1%
合計	355,400	

4.1　データセットの概要

教師データを作成するために，ヤフー株式会社が運営する総合映画情報サイトである Yahoo!映画に投稿された「作品ユーザーレビュー」を収集した．収集した 6 万件の映画に対するレビューには，点数だけを記録するためのものが多く含まれた．そこで，本文が空であったり，一言しかないものなどを除くために，本文が 10 文字以内のものを取り除いた．

本実験で対象としたユーザ数とそのユーザの投稿したレビュー数の関係を集計したものを表 1 に示す．

本実験では，モデルの学習のためには，同一ユーザによるレビューが複数必要になる．そのため，レビューを 50 件以上投稿したレビュアによる投稿だけを学習の対象とした．そもそもほとんどの利用者はたかだか数件しかレビューを書かないので，学習に用いたレビュアは，全レビュアの 1.32%程度になった．

4.2　提案手法の実装

提案手法を実際のデータに適用するために，学習済みの BERT モデルを，レビューサイトから収集したデータでファインチューニングした．事前学習モデルは東北大学の公開している既存のモデルを用いた．レビューデータは Yahoo!映画のものを用いた．

日本語の学習済みの BERT モデルとして，東北大学の乾研究室の公開しているモデルを使用した．これはモデルのアーキテクチャ，学習は元の BERT と同じ構成である．学習データには 2020 年 8 月 31 日時点の日本語版 Wikipedia を用いている．ファイルサイズは 4GB で，約 3 千万文が含まれている．また語彙サイズは 32,768 である．

レビューをある程度書きなれていて，自分の文体が確立していると考えられるユーザだけを対象に，モデルの学習を行う．そのために，レビューを 50 件以上書いているユーザを抽出し，同一ユーザ判定に用いた．こうしたユーザのうち，ランダムに抽出した 2,000 ユーザを教師データとして用いた．正例と負例を作成するために，同一のレビュアに書かれたレビューと，異なるレビュアによって書かれたレビューのペアを作成した．まず，1 ユーザあたりランダムに 50 件のレビューをランダムに抽

表 2　ファインチューニング時の BERT のパラメータ

パラメータ	値
入力の最大系列長	128
バッチサイズ	16
オプティマイザ	Adam
学習係数	2e-5
エポック数	5
精度	0.72

出した．正例として，その 50 件のレビュー中で，全ての組み合わせである 2,450 個のレビューのペアを作成した．負例として 50 件のレビューそれぞれにそのユーザ以外によって書かれたレビューである 50 件を組み合わせ 2,500 個を作成した．こうして作成した 4,950 個のレビューのペアについて，2,000 人分，合計で 9,900,000 個の学習データを作成し，実際に学習を行った．

4.3　学習時のパラメータ

学習時のパラメータ，学習後のテストデータによる精度を表 2 に示す．Sentence-BERT の入力層にあたる BERT モデルでは，最大 512 トークンを入力できるが，精度と学習速度の兼ね合いから，128 トークンに切り上げて学習を行った．BERT のバッチサイズは Liu ら [9] により，事前学習においてはより大きなバッチサイズで学習することが有効であると指摘されている．そのため学習時に使用した GPU の扱える上限であった 16 に設定した．エポック数は 5 としているが，検証誤差の値が更新されないとき学習の早期終了を行うよう設定した．

テストデータは教師データとして用いたユーザ外から作成し，精度は 0.72 であった．

4.4　比較手法の実装

提案手法の特性を評価するために，提案手法の他に 3 つの比較手法を用意した．すなわち，

- **提案手法**：本研究で提案している著者を推定するタスクでファインチューニングした BERT モデル，
- **BERT**：ファインチューニングを施さない，もとのままの BERT モデル，
- **Doc2Vec**：BERT 以前に一般的に用いられてきた分散表現化手法，
- **ランダム**：完全に無作為に抽出したレビュー

を比較することで，本研究におけるファインチューニングの性質を明らかにする．本節では各種法の詳細について述べる．全手法に共通して，入力に用いるテキストに関しては，提案手法と同様の前処理を行っている．

BERT は，ファインチューニングを行わない BERT 言語モデルによる比較手法である．提案手法のベースとして用いた東北大乾研究室が公開している学習済み日本語 BERT モデルをファインチューニングを行わずに用いた．推論時には提案手法と同様に入力されるテキストは 128 トークンまでに切り上げ，BERT モデルからの出力を Average-Pooling することで分散表現を得る．

表 3　学習済み Doc2Vec のパラメータ

パラメータ	値
dm	0
vector-size	300
windows	15
alpha	0.025
min-count	5
sample	1e-5
epochs	20
dbow-words	1

Doc2Vec はネット上に公開されている日本語コーパスによる学習済みモデルを用いた比較手法である．学習アルゴリズムには distributed bag of words (PV-DBOW) を用いている．Python の自然言語処理用のライブラリである Gensim を用いて実装されている．Gensim におけるモデルの主要な学習パラメータは表 3 に示す．学習元のコーパスとしては 2019 年 1 月 14 日時点での Wikipedia の CirrusSearch のダンプデータを用いている．

ランダムサンプリングは，それぞれのタスクがどれほど困難であるかを表すためのベースライン手法である．他手法ではその手法で得られた分散表現から類似度の高い 5 つのレビューを抽出しているのに対して，ここでは比較元となるレビューと同一のタイトルの映画に対するレビューの中からランダムに 5 つ抽出したものを比較対象のレビューとして用いる．

4.5　クラウドソーシングによる評価実験

提案手法の特性を評価するために，クラウドソーシングによる大規模な被験者実験を行った．実験はアンケート形式で，被験者は 2 つのレビュー文がどういった観点で類似しているかを回答した．被験者収集には，クラウドソーシングプラットフォームであるランサーズを利用した．

実際のクラウドソーシングサイトでのアンケートフォームのスクリーンショットを図 3 に示す．被験者には，比較元となる 1 つのレビュー文に対して，20 件の異なるレビューが提示された．これらのレビューは，同じ映画に対する，4 つの手法で類似度が高いと判定された，上位 5 件のレビューを織り交ぜたものである．被験者は，それぞれのレビューについて，比較元とどのような点で類似性があるか，8 つの質問について「はい」，「いいえ」の 2 択で回答した．実際の質問項目を，表 5 に示す．アンケートフォームの先頭には例として，より質問項目に関して判定しやすいレビューを作成し，その根拠とともに回答例を提示した．そこで提示した例を，図 2 に示す．

比較元となるレビューは映画 20 タイトルについてそれぞれ 5 件ずつ，合計 100 件のレビューを用いた．対象とする映画は，人手で選出した．この際，評価実験で各手法ごとに同程度の文字数のレビューを比較可能な，250 件以上レビューの付いた映画を対象にした．また，レビュー数の多い映画，少ない映画に偏りがないように，段階的にレビュー件数ごとに映画を選定した．邦画，洋画の割合，ジャンルなどについても，ばらつきがないように人手で調整した．実際に対象とした映画と，それぞ

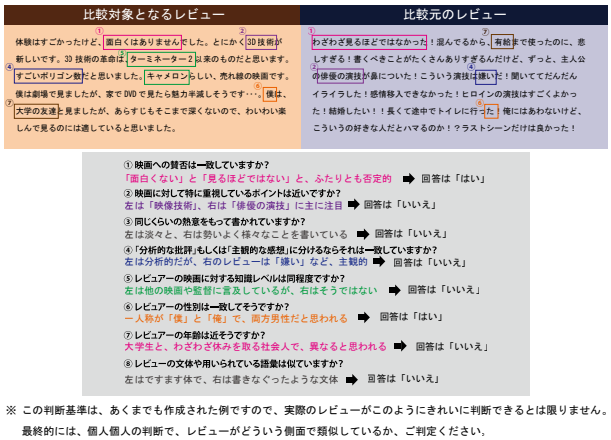

図 2　実際にクラウドのユーザに提示した回答例

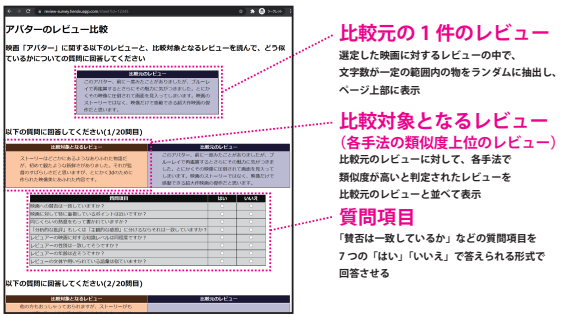

図 3　実際のオンラインクラウド評価時のスクリーンショット．2 件のレビューのペアに対して，8 種類の類似性に関する質問に回答する

れに対する有効なレビュー数を表 4 に示す．

レビューを比較する際に，レビューの文字数が 50 文字から 300 文字のものだけを対象とした．これは，提案手法，比較手法ともに，一定の分量のテキストが入力にあることを前提としているためである．一言の感想のようなレビューを取り除くために，50 文字以内のレビューについては全て除去した．また文章量が比較対象と大きく開きがある場合，正確な回答が困難であるとため，文字数が 300 文字を超えるものは除去した．

各レビューについての質問は回答者の上限が 3 人になるように提示した．クラウドソーシングサイトで実験を行う上での制約として，不誠実なクラウドワーカを判別して除外する必要があった．そのために，質問項目のうち指定のものにのみ「はい」と回答するように促す内容の，ダミーのレビューを質問中に 1 件，紛れ込ませた．その指示に従わなかった回答者の回答は，信頼性の低いデータであるとして，除去した．

4.6　実験結果

300 件のタスクから，234 件の有効な回答が得られた．具体的な推薦結果の例として「インターステラー」の映画レビューを比較元，提案手法による上位 1 位のもの，BERT による上位 1 位のものを表 6 に示す．

クラウドソーシングによる回答結果を表 5 に示す．結果については t 検定を行い，提案手法と BERT による手法を比較して，有意差が認められた項目については数値に*をつけて表記している．提案手法と，ファインチューニングを行っていない BERT による比較では，「レビューの文体や用いられている語彙は似ていますか？」という質問項目に関しては有意差が認められた．このことから同一ユーザの判定タスクによって BERT をファインチューニングすることによって，文体や用いられている語彙に敏感な分散表現化が行えることがわかった．他の質問項目では提案手法による有意差は認められなかった．

表 4　評価実験に用いた映画タイトルと対象としたレビュー件数

映画名	レビュー数
シン・ゴジラ	5,850
アバター	4,350
ALWAYS　三丁目の夕日	3,160
アナと雪の女王	2,742
ラ・ラ・ランド	2,466
インターステラー	1,942
ダ・ヴィンチ・コード	1,863
思い出のマーニー	1,729
パシフィック・リム	1,708
ショーシャンクの空に	1,639
万引き家族	1,449
ベイマックス	1,388
図書館戦争	1,090
重力ピエロ	850
紙の月	665
ヘイトフル・エイト	367
フェイス／オフ	285
コンタクト	261
若おかみは小学生！	254
ロード・オブ・ウォー	253

ランダムサンプリングと，他手法のうち，値が最小のものとの比較では，「映画に対して特に重視しているポイントは近いですか？」，「同じくらいの熱意をもって書かれていますか？」，「「分析的な批評」もしくは「主観的な感想」に分けるならば，それは一致していますか？」，「レビュアーの映画に対する知識レベルは同程度ですか？」，「レビュアーの年齢は近そうですか？」，「レビューの文体や用いられている語彙は似ていますか？」という質問項目に関しては有意差が認められた．

5　考　　察

結果から，同一ユーザの判定タスクによって BERT をファインチューニングすることによって，文体や用いられている語彙に敏感な分散表現化が行えることがわかった．あるユーザによって書かれたレビュー文を元に，そのレビューが対象としている映画のレビューを推薦するという設定においては，クラウドソーシングでのアンケートの結果からは提案手法の有用性を示すことはできなかった．本章ではいくつかの観点に注目して提案手法によって得られる分散表現の特徴について検討し，論じる．

はじめに，文体と性別，年齢の関係について注目する．提案

表 5　クラウドソーシングサイトで実際に提示した質問項目と，各手法に「はい」と答えた割合
（提案手法に対し BERT，片側 t 検定における p 値:* < 0.05）

質問文	提案手法	BERT	Doc2Vec	ランダム
映画への賛否は一致していますか？	0.59	0.64	0.63	0.56
映画に対して特に重視しているポイントは近いですか？	0.31*	0.39	0.43	0.22
同じくらいの熱意をもって書かれていますか？	0.47	0.45	0.42	0.33
「分析的な批評」もしくは「主観的な感想」に分けるならば，それは一致していますか？	0.51	0.54	0.55	0.40
レビュアーの映画に対する知識レベルは同程度ですか？	0.48	0.51	0.51	0.36
レビュアーの性別は一致していそうですか？	0.59	0.62	0.58	0.55
レビュアーの年齢は近そうですか？	0.53	0.53	0.47	0.40
レビューの文体や用いられている語彙は似ていますか？	0.48*	0.42	0.41	0.28

表 6 「インターステラー」の推薦結果の例

手法	レビュー文
比較元	アインシュタインの相対性理論を理解した上で鑑賞なぞできなかったのだが，ワームホールを抜ける映像や，わずか数分の出来事が，地球時間では数年単位で進んでしまうという現象を，観客側としては俯瞰で観ることができたし，よくできた SF を見せてもらった気がする．そして 5 次元世界の映像化．映像作家ならば腕の見せ所だろうが，キューブリックが見たとしたら一体どのように評しただろうか?個人的には大いに楽しませてもらったが，この尺でも全く苦にならず楽しめたのは，この想像力豊かな映像たちのおかげ．いずれまた監督には驚かせてもらいたい．
提案手法	SF(サイエンスフィクション) が文字通り科学的な空想に基づいたフィクションであるとするなら，この作品は間違いなく最先端の sf 作品と言っていい．少なくとも今後作られるであろう SF 映画に平面のワームホールや，ただの黒い円として描かれたブラックホールが登場する機会は激減するだろう．インターステラーの製作には世界的な物理学者が関わり，脚本家は相対性理論を一から学んで製作に挑んだ，裏付けのなされた SF 作品だ．もちろん理論的に描いた小説はこれまで数多くあったが，難解な題材をノーラン監督は映像化し，更に娯楽作にまで見事に昇華させている．近年稀にみる SF の傑作だ
BERT	相対性理論のよくわからない論理を字幕だけで理解するのはほとんど不可能なので，頭の中は「なんでそうなるの?」ばかり．それらが演技や映像の素晴らしさを邪魔している気がしました．むしろ『2001 年宇宙の旅』のように，思い切り哲学的な作品に振り，鑑賞者に考える余地を残すのも一興かと思いつつ，しかし興業を考えるとそうもいかないんでしょうね．テレンス・マリックがこの作品を手がけていたら，どんなことになっていたのでしょうか (笑).

手法と BERT による手法を比較すると，文体については有意差が認められたものの，性別，年齢の質問項目においては有意差が認められなかった．このことから，提案手法において仮定していた，文体や用いられている語彙が近いユーザであれば性別や年齢などの属性も近いはずであるということが正しいとは言えないということがわかった．ただし， 図 2 で示した，被験者に提示した例では簡単のため「ですます」調であるか否かを判断の基準として示していた．そのため被験者側の類似性の判断の基準がこのような表層的な部分での文体を基準にしてしまっていて，性別，年齢に繋がるような意味での基準から離れてしまっていたという可能性も考えられる．

次に,「注目している観点」の質問項目に対する結果について議論する．提案手法では，レビュー推薦において特に重視するべきと思われる,「重視しているポイント」の近さの項目では BERT や，より軽量な Doc2Vec にも劣るという結果になった．これは「重視しているポイント」の抽出には，高度な文脈情報よりも「俳優」,「映像」といった具体的な単語が重要になるからだと考えられる．

今回の研究では，実際に複数のレビューを書かせて，レビューを推薦するというタスクでの評価実験が不可能であった．そのために，クラウドソーシングサイトにおけるアンケートをもとに，間接的な評価を行わざるを得なかった．このような実験上の制約から生じた，十分に分析できなかった要素について論じる．

そもそも，1 問 1 答形式では，本当に細かい類似度に関する比較が行えなかった可能性がある．推薦結果の例として「インターステラー」の映画レビューを表 6 に示した．これらの提案手法と BERT によって選ばれたレビューは「注目している観点が近いか」という質問項目について，両者ともに「はい」と回答された．これら 2 件のレビューは，ともに「相対性理論を正しく映像化している」ことに言及している．そのため回答者は,「はい」を選択しがちである．しかし BERT の結果を見ると,「それらが演技や映像の素晴らしさを邪魔している」という記述がある．このことから，レビューの著者について深く考えると，映像や演技を重視している可能性もある．今回の質問では，被験者は，あくまでもレビュー文中にある観点に対する言及が含まれているかで判定を下しがちであった．一方で，実際のレビュー推薦というアプリケーションを考えると，比較元の著者に対しては提案手法のレビューを提示する方が有効な可能性がある．このように，映画などの感性が関わるレビューにおいては，明示的なフィードバックが与えられないことが多い．そのためオフラインテストを行うことができず，また当人でな

い限り評価が曖昧である．正確な評価を行うためには，A/B テストなどの評価手法が必要になると考えられる．

最後に，ランダムサンプリングと他手法について比較する．ランダムサンプリングによる推薦結果はすべての質問項目において，他の手法に大きく劣ることがわかった．このことから，レビューの提示順序が投稿順である多くのレビューサイトにおいては，Doc2Vec を使った軽量な手法であっても，導入することによってユーザエクスペリエンスを向上させることができる余地があるといえる．

6 まとめと今後の課題

本研究では，レビューの著者にファインチューニングした BERT を用いることで，レビュー推薦を行う手法について提案した．提案手法ではある 2 つの任意の映画に対するレビュー文を与えた際に，それらのレビューの著者が同一であるかを推定するタスクでファインチューニングを行った．提案手法と他比較手法による推薦結果を大規模なクラウドによるアンケートを行うことによって手法の有用性について議論した．

ユーザが未だレビューを書いていない映画に関するレビューの推薦や，映画自体の推薦においては，考察で記した直接的な具体的な単語による比較ができない．そのためこのような場合，提案手法である文体や用いられる語彙に敏感な分散表現は比較手法に比べ有用である可能性があると考えられる．具体的にはここで得られる同一ユーザによるレビューの分散表現を用いてユーザの分散表現を作成し，そのユーザ間の類似度の近いユーザのレビューを推薦することや，その類似度を協調フィルタリングにおけるユーザ間類似度に転用することによって映画自体の推薦を行うなどである．これらの実装について今後の課題としたい．

謝　　辞

本研究は JSPS 科研費 18K18161（代表:荘司慶行），21H03775（代表：大島裕明）の助成を受けたものです．ここに記して謝意を表します

文　　献

[1] Oren Barkan and Noam Koenigstein. Item2vec: neural item embedding for collaborative filtering. In *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6. IEEE, 2016.

[2] Van-Thuy Phi Liu Chen and Yu Hirate. Distributed representation-based recommender systems in e-commerce. In *DEIM Forum 2016 C8-1*, pp. 1–6. DEIM, 2016.

[3] Gaojun Liu and Xingyu Wu. Using collaborative filtering algorithms combined with doc2vec for movie recommendation. In *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, pp. 1461–1464. IEEE, 2019.

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[5] Nils Reimers, Iryna Gurevych, Nils Reimers, Iryna Gurevych, Nandan Thakur, Nils Reimers, Johannes Daxenberger, Iryna Gurevych, Nils Reimers, Iryna Gurevych, et al. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2019.

[6] Qing Guo, Wei Feng, Ce Zhou, Rui Huang, Liang Wan, and Song Wang. Learning dynamic siamese network for visual object tracking. In *Proceedings of the IEEE international conference on computer vision*, pp. 1763–1771, 2017.

[7] Gustavo Penha and Claudia Hauff. What does BERT know about books, movies and music? Probing BERT for conversational recommendation. In *Fourteenth ACM Conference on Recommender Systems*, pp. 388–397, 2020.

[8] Itzik Malkiel, Oren Barkan, Avi Caciularu, Noam Razin, Ori Katz, and Noam Koenigstein. RecoBERT: A catalog language model for text-based recommendations. *arXiv preprint arXiv:2009.13292*, 2020.

[9] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.