

大規模アクセスログを用いた映画視聴行動を促すレビューの特徴分析

伊藤 翔[†] 莊司 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

^{††} 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: [†]kakeru@sw.it.aoyama.ac.jp, ^{††}shoji@it.aoyama.ac.jp, ^{†††}sufujita@yahoo-corp.jp,

^{††††}duerst@it.aoyama.ac.jp

あらまし 本論文では、実サービスのアクセスログから、実際に読み手の行動を誘発したレビューの特徴を分析する。Yahoo! JAPAN の「Yahoo!映画」サービスでは、毎日数多くの利用者が映画レビューを読み、その後、配信サイトで映画を視聴している。そこで、アクセス履歴から個別のレビューにアクセスした直後に配信サイトにアクセスしているログを抽出することで、実際に視聴行動を起こさせたレビューを特定した。このようなレビューを映画視聴行動を促すレビューであると仮定し、任意のレビューに対して、視聴行動を促すかを判定するランダムフォレスト分類器を学習した。60日分以上に及ぶ大規模な実データを用いた交差検証により、各特徴の重要度を算出した。これにより、どのような内容、投稿者、対象を持つレビューが読者の映画視聴を促すか明らかにした。

キーワード 行動ログ分析, オンラインレビュー

1 はじめに

2020年までに公開された映画の総数は、53万本を超える¹。仮に1本の映画が2時間だとすると、すべての映画を見るためには、120年間の視聴時間を要する計算になる。そのため、現代のすべての人類は、人生でどの映画を見てどの映画をあきらめるか、短時間で判断しながら生きているといえよう。

オンラインレビューサイトは、こういった日常的な取捨選択の判断を助ける、きわめて重要な情報源である。今日ではショッピングサイトの商品や、動画配信サイトの映像作品、地図上に配置された施設をはじめ、学校、医者、個人など、ありとあらゆるものに対するレビューが、インターネット上に溢れかえっている。そして、インターネット上の評判を確認することによる意思決定は、もはや生活の一部になってきている。

一方で、これらのレビューサイトに投稿されたレビューのうち、すべてが意思決定を行う上で有用であるとは限らない。ウェブ上には様々な人が投稿する多様なレビューが存在し、その質も玉石混交である。レビューサイトから、効率的に意思決定に用いることのできる情報だけを選択することは、困難である。現在では、多くの商品に対し、きわめて多くのレビューが存在する。この際、意思決定を行うために多くのレビューを読み進める行為は、大きな時間的コストを要することが知られている [1]。

ここで、意思決定に役に立つレビューとそうでないレビューを分析することで、レビューを「意思決定に役立つ順」でランキングする方法について考える。これまでにレビューの有用性に関する研究は盛んに行われてきており、どのような特徴を持つレビューが有用かが議論されている。しかし、一般的にこれらの研究では、あるレビューに対して読者が「このレビュー

は役に立った」という投票ボタンを押したかどうかを、そのレビューの有用性として用いている。こういった読者投票による有用性が、実際の意思決定に用いることができるかと一致するとは限らない。どのような時に読者が「役に立った」あるいは「役に立たなかった」ボタンを押すかを考える。例えば、工作行為として自分の支持する商品への高評価レビューに「役に立った」ボタンを押すかもしれないし、単に読者本人の意見と似た意見に対して「役に立った」ボタンを押すかもしれない。こういった投票行為は、一般的な利用者からすると、意思決定において有用とは限らない。そこで、我々は、本当に役に立ったレビューを読んだ場合には、「役に立った」ボタンを押すより先に、その商品を購入するなり、行動に移していると考えた。

本研究では、レビューを意思決定に有用な度合いで順序付けするために、実際の大規模ログデータを分析した。はじめに、Yahoo! JAPAN の「Yahoo!映画」において、レビューを読んだ直後に、配信サイトで実際に映画を視聴している行動ログを収集した。次に、この際に読まれていたレビューを「視聴行動を誘発するレビュー」としてラベル付けし、そうでないレビューと分類する分類器を学習した。分類の際に、

- 文体や使っている語彙など、レビューの内在的な特徴、
- そのレビューの著者のレビュー投稿本数などのレビューに付随する特徴、および
- 他に何件レビューがついているかなどの、そのレビューの対象となるアイテムに付随する特徴

を、それぞれレビューを表す特徴量として用いた。このような学習を行った分類器について、用いた特徴の寄与度を分析することで、読者の行動を誘発するレビューの特徴を明らかにする。

このように抽出した行動を誘発するレビューの特徴を持つレビューについて、被験者評価によって、意思決定に有用であるかを評価した。

以下に本論文の構成を示す。本章では、研究の背景と目的に

1: <https://www.imdb.com/pressroom/stats/>

ついて述べた。第2章では関連研究として、目的や仕様技術の類似した研究分野について、それぞれ紹介して論じる。第3章では本研究で提案する手法を述べる。第4章では実際の「Yahoo! 映画」のアクセスログを使った分析について述べる。第5章で被験者実験とその結果について述べ、第6章で結果について考察する。第7章でまとめと今後の展望について述べる。

2 関連研究

本研究は、レビューの様々な特徴を用いて視聴行動を促すレビューを特定する。そこで、レビューに代表される口コミ情報(eWOM: Electric Word of Mouse)を対象とし、レビューの特徴および、その書き手や対象となるアイテムとレビューの特徴の関係に触れた関連研究について紹介し、本研究との差異や位置づけについて論じる。

2.1 有用なレビューの分析

レビューの特徴と有用性の関係について、現在までに多くの研究が行われてきている。Zhou [2]らは、レビューの有用性の高さについて、レビューを読むことで製品を評価する能力が向上するものと定義している。この際、レビューが持つ特徴を数値的特徴と文章的特徴の2種類に分類し、相互に影響を及ぼすという仮定で検討している。数値的特徴として、文章のおおよその長さや、星の数などで表される評価値の高さなど、文章の内容を読む前に読者が得られる情報を用いている。また、文章的特徴として、文章の感情極性や、一文に含まれる平均単語数など、文章の内容を読むことで読者が得られる情報を用いている。

Zhouら[2]は、文章の長いレビューほど、有用であると認識されやすいことを示している。Zhouらはこの結果について、レビューの読み手が、書き手の有用性の提供意思がレビュー文章の長さとして現れていると考え、レビュー文章が持つ有用性判断への影響を与えたと考察している。これは、レビューの読み手が、書き手の有用性の提供意思がレビュー文章の長さとして現れているためだと考えられる。

数値的特徴のうち、評点と有用性に関する研究は多く行われてきたが、研究によって、相反する結論が導かれている。Panら[3]は、評点が高いレビューほど、有用なレビューとして認識されやすいことを述べている。しかし、Chuaら[4]は、評点の高いレビューほど、有用なレビューとして認識され難いと述べている。Mudambiら[5]は、これらの二つの結果とは異なり、評点の有用性に非線形な関係が存在することを示した。評点が高いレビューおよび、低いレビューは有用性があると認識され難く、平均的な評点を持つレビューが最も有用と判断されやすいことを述べた。

ほかの文章的特徴としてレビューの本文が肯定的か否定的かという極性が消費者の購買行動に及ぼす影響に関しても、研究がおこなわれている[6,7]。しかし、YuanYuan[8]らは、これらの結果が一貫したものではないことを述べている。

Hongら[9]は、レビューの有用性に影響を与える特徴に関

する研究[10,11]のメタ分析において、レビュー文の複雑さとレビューの投稿時からの経過時間の特徴量が、有用性に関して正の相関があると述べた。

同様にレビューの有用性に着目した研究では、Chenら[12]が他のユーザから「役に立った」という評価を得られているレビューほど、読み手の購買意欲に対して強い影響を持つことを述べている。しかし、大部分のレビューは他のユーザからの評価が不足し、「役に立った」という指標が有用性評価に充分でないことが指摘されている[13,14]。理由として、新しく投稿されたレビューは、他のユーザから評価を受けるのに十分な時間がない[13]ことが考えられている。

2.2 レビューの質の分析

レビュー本体でなく、その著者がレビューの有用性に与える影響についても議論されている。Connors[15]らは、レビューの有用性判断において、読み手が専門家によって書かれたと判断可能なレビューは、有用性が高いと判断される傾向があることを論じた。レビューの書き手の情報開示の度合いと有用性に関する研究[16-18]では、開示の度合いと有用性に関して正の相関があることを述べている。同研究で使用されたデータセットでは、レビューの書き手が自分の情報をユーザ情報として登録することが可能だった。しかし、本研究で用いるデータセットでは、ユーザは任意の情報を登録すること不可能であり、レビューの読み手は、書き手に関して、投稿レビュー数や、他の投稿レビューの内容などサービス上での客観的なパラメータ以外を取得不可能である。そのため、レビューの読み手が、書き手の専門性を判断する機会や、傾向に差が生じる可能性がある。

2.3 対象によるレビューの変化の分析

本研究では、レビューの対象とされるアイテムが持つ特徴を学習に用いる。Nelsonら[19]は、eWOMを対象とした調査において、製品のタイプを検索商品と体験商品の2種類に分類する手法を提案している。Mudambiら[5]は、製品のタイプがレビューの読み手による有用性認識に影響を与えていることを述べている。Huangら[20]らは、体験商品に対するレビューにおいて、客観的な情報で構成されるレビューよりも、主観的な感想や印象で構成されるレビューの方が有用と判断されやすいことを述べている。

Luanら[21]は、視線追跡装置を用いた実験を行い、被験者によるアンケートを用いた実験の結果と比較を行い、Huangら[20]の示した結果を検証した。同研究では、ユーザの視行動は商品のタイプのみによって変化するわけではなく、商品のタイプとレビューの組み合わせによって影響を受けることを示した。先行研究で述べられた結果との比較によって、ユーザが体験商品の評判を閲覧する際、レビューの種類に関係なく、多くの情報を得る目的で行動すると考察している。

本研究は映画のレビューを対象とする研究であり、関連の研究において、体験商品のレビューで重要とされた特徴群と同様の特徴が、影響力を持つ可能性がある。

3 提案手法

ある1件のレビューが与えられた際に、そのレビューが人の行動を誘発するかどうかを推定する、ランダムフォレスト分類器を学習する。手法全体の流れを、図1に図示する。実際の分析では映画のデータセットを用いるので、映画を例にとって説明する。

ある1件のレビューは、大別して3種類の特徴によって表現することができる。ひとつ目の特徴は、レビューの対象となったアイテムの持つ特徴である。例えば、映画のカテゴリであったり、監督の有名さであったり、その映画がどれほど人気かなどが特徴として用いられる。このような特徴が必要になる過程として、アイテムの性質によって、レビューの使われ方が異なる可能性があるためである。例えば、アクション映画であれば簡潔なレビューが読者に「見たい」と思わせるかもしれないし、特定の個性派監督の映画の場合は評点の低いレビューの方が逆に「見たい」と思わせるかもしれない。

二つ目の特徴は、そのレビュー単体の持つ内容そのものである。例えば、レビューの長さは、そのレビューが意思決定に役立つかの重要な指標の一つであると考えられる。他にも、監督名や女優名を含んでいるかや、オノマトペや感情語の含まれている割合は、そのレビューを読んだ際の印象に直結する。また、レビュー全体で扱っているトピックも、重要な要素のひとつであると考えられる。

三つ目の特徴は、そのレビューを投稿したレビューアを表す特徴である。レビューアのプロファイルや過去の投稿傾向は、レビューの質と関係するほか、レビュー読者の意思決定に影響する可能性がある。例えば、ほかの人と同じような語彙を使っているにもかかわらず、人を行動に駆り立てる煽情的なレビューを書くユーザがいるかもしれない。また、極端な例では、ファンを多く持つレビューアもいるかもしれない。そこで、レビューアのこれまでのレビュー本数などを特徴として扱う。

このような特徴から、その1件のレビューが人の行動を誘発するか、分類器で分類する。この際、正解ラベルは、実際の行動ログから作成する。この手法では、ある実際の行動を伴うサービス（今回の場合は、映画視聴を伴う配信サイトの配信ページ）へのアクセスログと、実際の個別のレビューページへのアクセスログの両者を紐づけて用いる。

はじめに、動画配信サービスと、レビューサイトの両方にアクセスしたことがあるユーザのアクセスログを収集する。次に、レビューを読んだのちに配信サービスに移動しているログを、「実際に行動が誘発されたログ」として抽出する。これらの「実際に行動が誘発されたログ」について、レビューごとにまとめて集計する。1件のレビューに対して、そのレビュー単体にアクセスした人数と、実際に映画配信サイトに移動したユーザの比率を取り、一定以上高かったレビューを「行動を誘発するレビュー」とみなす。

このように学習した分類器は、実用上の2つの利点を生み出す。ひとつ目は、未知のレビューが与えられた際に、そのレ

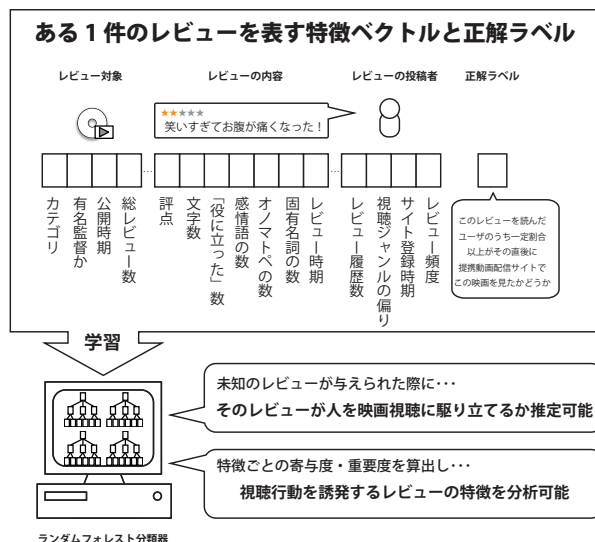


図1 提案手法の全体の流れ。与えられた1件のレビューに対して、ランダムフォレスト分類器を用いて「人の視聴行動を誘発するレビューであるか」を分類する。

ビューがどれだけ人の行動を誘発できそうか推測することである。二つ目は、特徴量の分類への寄与度から、どのようなレビューが人の行動を動かすか分析することである。

4 実データを用いた分析

本研究では、Yahoo!映画のレビューデータおよび、Yahoo!映画のアクセスログを用いて、実際にユーザの意思決定と行動を推定することにより、意思決定に寄与するレビューを特定した。

4.1 ログデータのクレンジング

分析に用いるデータとして、映画配信サイトGYAO!で配信されている映画へのアクセスログと、それらに対するレビューのレビュー内容データと、それぞれのレビューへのアクセスログを用いた。アクセスログは、2019年11月1日から2019年12月31日までの、61日分を用いた。アクセスログは、ユーザと映画とレビューと訪問日時を一つのセットとしたレビュー閲覧ログデータの集合と、ユーザと映画と訪問日時を一つのセットとした映画視聴ログデータの集合に別れる。

Yahoo!映画は、映画配信サービスであるGYAO!²と連携している。そのため、Yahoo!映画の映画情報ページには、GYAO!で視聴可能な映画の場合、配信ページへのリンクが張られている。そこで、本研究ではYahoo!映画内から、GYAO!の映画視聴ページへ遷移したログを抽出することで、ユーザが実際に映画を視聴した行動ログだけを抽出することができた。

このように、映画視聴ログデータから、あるユーザがある映画を観たというデータを抽出した。その後、レビュー閲覧ログデータから、同一ユーザの同映画に関するレビューに対する閲覧記録を取得する。これらの閲覧記録の中から、映画の視聴の

2: <https://gyao.yahoo.co.jp/>

表 1 レビューがユーザに視聴行動を促す度合いを表すスコアに使用する記号の定義

記号	定義
N	スコア算出の対象とするレビューの数
a_i	レビュー i が閲覧された後に映画の視聴があった件数
b_i	レビュー i が閲覧された後に映画の視聴が無かった件数
$induce_i$	レビュー i が映画の視聴を促す度合い

前に閲覧されていたレビューを、映画の視聴を促したレビューであると仮定した。

ここで、データセットから、一度も GYAO! を使ったことのないユーザは取り除いた。これは、Yahoo! 映画を利用するユーザのうち、GYAO! 以外の映画配信サイトで映画を視聴するユーザや、そもそもインターネット上で動画を視聴しないユーザが存在すると考えられるためである。本データセットでは、GYAO! で映画視聴を行うユーザと、それ以外のユーザに存在する特徴差が学習されることを防ぐため、GYAO! で映画を視聴した実績があるユーザのみを対象として、正解ラベルを付与した。

4.2 正解ラベルの付与

クレンジングしたログデータについて、レビューがユーザの視聴行動を誘発する度合いを、そのレビューの閲覧数と行動を引き起こした回数の比率から計算する。表 1 に記号と定義を示す。誘発度合いの算出の対象とするレビューの数を N とする。このとき、任意のレビュー i について、閲覧された後にレビューの紐づく同映画の視聴があった件数を a_i とする、また、閲覧された後にレビューの紐づく同映画の視聴が無かった件数を b_i とする。このとき、レビュー i の視聴行動を誘発する度合い $induce_i$ を

$$induce_i = \frac{a_i}{\sum_{k=1}^n a_k} - \frac{b_i}{\sum_{k=1}^n b_k}$$

と定義する。

実際の分析では、ログから抽出した 70,000 件のレビューに対して、実際の視聴行動を起こした割合 $induce_i$ を算出した。70,000 件のレビューに対して、スコアが高いものから 7,000 件を正解データとし、それ以外のレビューから 7,000 件を不正解データとしてランダムに抽出した。

4.3 レビューに内在される特徴量の計算

正解ラベルが付与されたレビュー 14,000 件について、それぞれを複数次元からなる特徴量として表現する。使用した特徴を表 2 に表す。

1 件のレビューを特徴量として表現するために、はじめに、レビューそのものに含まれる内在的な特徴に注目する。レビューの情報量に注目した特徴として、レビューの文字数、タイトルの文字数と、ネタバレフラグを用いた。ネタバレフラグは、そのレビューに物語の核心が含まれているかを 0 か 1 で表したものである。ネタバレを含む投稿は、ウェブサイト上では、クリックするまで本文が表示されない。

レビューの読みやすさも、レビューが人の行為を誘発するか大きく影響すると考えられる [5]。そこで、可読性の指標として、1 文の長さ（すなわち、句点や感嘆符などで区切られた部分の平均長）、1 行の文字数をそれぞれ特徴として用いた。

レビューの内容を表す特徴として、印象タグ、レビュー本文に含まれる語のトピック、総合評点、個別の観点に基づく評点をそれぞれ用いた。印象タグはレビューを投稿する際にレビューアが「イメージワード」としてタグを付与できる機能である。具体的には、「泣ける」や「ファンタジー」、「かっこいい」などの 20 種類から、任意の個数のタグを選択して付与する。レビュー本文がどのような語を使っているかを表すために、LDA を用いて本文中の語を 50 次元からなるトピックに変換した。この際、レビューに含まれる語を辞書的に分類するために、あらかじめデータセットのレビュー文で学習したモデルを用いた。実際の特徴の分析のために、特徴ベクトルの大部分を占めるトピックの影響を、他の特徴と合わせる必要があった。そのために、あるレビュー文は 1 つのトピックに属するという仮定を置き、レビュー内容を one-hot ベクトルとして特徴化した。レビューの評点も、レビューの内容を表す重要な特徴である。多くのレビューサイトにおいて、評点は星の数（すなわち、星ひとつから星 5 つ）として表される。一部のレビューサイトでは、映画に対して総合的な評点のほかに、個別の観点から評点を与えることができる。今回用いたデータでは、対象の映画に対して、「物語」、「配役」、「演出」、「映像」、「音楽」の 5 つの観点ごとに、5 段階で観点ごとの評点を扱うことができた。これらの総合評点と観点ごとの評点を、それぞれ 1 次元ずつの特徴として扱った。

ほかに、そのレビューがどの程度人目に触れたかを特徴として含めた。具体的には、レビュー 1 件のページの被閲覧数と、レビューが役に立ったと感じた際に閲覧者が投票できる「役にち度」を、それぞれ特徴として用いた。

4.4 アイテムに付随した特徴量の計算

次に、レビューの対象となるアイテムごとの特徴として、メタデータ、映画の内容、映画の人気度、評判をそれぞれ特徴として用いた。そのレビューの対象としている映画の性質そのものが、閲覧者の視聴行動を促すかに大きく寄与すると考えられる。映画の持つ一般的な情報として、公開年代、タイトル長をメタデータとして用いた。

また、映画の内容や質を表す特徴として、同監督の撮影した映画の本数を、ベテラン監督による映画かどうかの指標として特徴に含めた。また、それぞれの映画は「アクション」「ファミリー」など十数種のカテゴリのうち一つ以上に紐づけられているため、特徴として用いた。各映画は映倫区分（すなわち、全年齢向け、18 歳以上を対象などの対象年齢）を持つため、これも特徴として用いた。

映画そのものの人気度も、重要な特徴である。人気のある映画は多くのレビューが集まると考えられるので、レビュー投稿数を特徴として用いた。ほかに、閲覧者が作成できる「見た映画」のリストに登録されている数、「見た映画」のリストに登

録されている本数を、それぞれ用いた。

映画の総合的な評判も、その映画を視聴するかの判断に寄与すると考えられる。そこで、ある映画に対するレビューの評点を集計し、平均をとったものを映画の総合評点とみなし、特徴として用いた。

4.5 レビュー投稿者を表す特徴量の計算

ある1件のレビューを表す特徴の一つとして、そのレビューを投稿したレビューアの特徴も考慮する。この際、レビュー個人を表す特徴として、年齢や性別などのレビューアのプロフィールは用いずに、レビュー行為そのものに関する特徴のみを用いた。これは、多くのレビューサイトにおいてプロフィールの入力は任意であり、また内容の信頼性も検証されていないためである。加えて、個人情報を使わず公開されているレビュー情報のみを用いることで、モデルがより多くのレビューデータで一般的に利用可能になる。

はじめに、あるレビューアがそのレビューサイトに投稿した経験をどの程度持つかが、レビューアのレビュー能力に関連すると考えられる [22]。そのために、レビューアの貢献度を特徴として用いた。具体的な特徴量として、投稿したレビューの情報量として、これまでのレビューの総投稿数、総文字数を用いた。また、このレビューアがどの程度支持されているかに関する特徴として、このレビューアの書いた各レビューの被閲覧数の累積と平均、これまでに獲得した「役立ち度」の総和をそれぞれ用いた。

そのほかのレビューアの特徴として、そのレビューアの各レビューアがどの程度の情報量を持つかを特徴として用いた。具体的には、全レビュー中でネタバレタグの付いたレビューを投稿した比率と、投稿したレビューの平均文字数をそれぞれ特徴として用いた。

最後に、レビューアが映画にどの程度詳しいかを表す指標として、観点評価差の平均と、印象タグ数の平均、獲得した役立ち度の平均をそれぞれ用いた。観点評価差の平均は、そのレビューアが観点ごとに評点をつける際に、特定の観点のみにこだわっていないか、星ひとつから星5つまでを幅広く使いこなしているかを表す特徴である。具体的には、各レビューでの観点ごとの最大評点と最低評点の差をとり、それをレビューアごとに平均した。印象タグ数平均は、映画に対して任意の個数付与できる印象タグについて、そのレビューアがこれまでに平均で何個タグ付けしたかである。ほかに、書いたレビュー件数が少なくとも、優れた小数のレビューだけを投稿する専門家のようなレビューアが存在する可能性がある。そこで、そのレビューアの書いたレビューが獲得した役立ち度投票の累計を、レビュー数で割ったものを、特徴として扱う。

4.6 ランダムフォレストによる分類と重要度計算

抽出した14,000件のレビューについて、そのレビューが実際に視聴行動を引き起こしたかの二値分類を行うランダムフォレスト分類器を学習した。この際、「レビューに内在する特徴と映画に付随する特徴と著者の特徴を学習した分類器」、「レビューに内在する特徴と映画に付随する特徴を学習した分類器」、「レ

ビューに内在する特徴を学習した分類器」の3種類の分類器を作成した。

作成した特徴量から、正解ラベルを推定するようランダムフォレスト分類器を学習した。ランダムフォレストの実装として Python の機械学習ライブラリである scikit-learn の実装を用いた。学習時のパラメータとして、木の本数を300本、木の最大深さを10に設定した。ほかのパラメータは、scikit-learn のデフォルト値を用いた。

はじめに、分類器同士の比較を行った。それぞれ学習機の F_1 スコアを表3に示す。最も精度が高かったのはレビュー内在特徴とアイテム付随特徴の2つの特徴を用いた分類器で、逆にすべての特徴を使ったモデルでは精度の低下が見られた。このことから、視聴行動を促したレビューの分類精度に、映画に付随する特徴が大きく影響していることが分かった。

次に、レビューに内在する特徴と映画に付随する特徴を学習した分類器の、特徴ごとの重要度を注目要素ごとに集計した。特徴ごとの重要度を表4にまとめる。それぞれ特徴の重要度から見ても、映画に付随する特徴が最も重要度が高く、合計0.42となった。次いでレビュー内容の重要度が0.30と高く、レビューアに関する特徴は0.27と最も低かった。最も重要度の高かった項目から、映画の人気度がユーザの視聴行動と特に大きく関係することが分かった。ほかに重要度の高かった特徴として、レビューアのレビュー経験の多さやレビュー内容が、ユーザの視聴を誘発するかと関係することが分かった。

5 評価実験

実際の分析で得られた知見が実世界で役に立つかを確認するために、被験者実験を行った。この実験では、ランダムフォレストによる分類結果が、実際の人間の認知と差異がないかを確認する。そのために、学習で用いなかった未知のレビューに対して、分類器でそのレビューが人の行動を誘発するかの推定を行った。

5.1 実験タスク

分類器に「視聴行動を促すだろう」と予測されたレビューと、「視聴行動を促さないだろう」と予測されたレビューを、それぞれ用意した。これらのレビューをランダムに被験者に閲覧させ、そのレビューを見て実際に対象となる映画が見たくなったかを、4段階で答えさせた。

実験の対象となる映画について、2019年の日本における興行収入の高い順に10本の映画を選んだ。各映画に対して投稿された全てのレビューに対して、分類器で推定を行った。この際、ランダムフォレストの分類の確信度に基づいて、実験に用いるレビューを抽出した。具体的には、「視聴行動を誘発する」「視聴行動を誘発しない」と分類されたレビュー群のうち、それぞれ分類の確信度が最も高かったレビューを5件ずつ抽出した。

被験者は、大学生6名である。はじめに、被験者に対して実験の対象となる映画の一覧を提示した。被験者は、映画に対して持っている事前知識から、対象の映画を観たい映画であるか

表 2 学習に用いた特徴

仮説	注目要素	特徴	次元数	
レビュー内在特徴	情報量	文章の長さ	1	
		タイトルの長さ	1	
		ネタバレフラグ	1	
	読みやすさ	文の長さの平均	1	
		改行の多さ	1	
		印象タグ	20	
	内容	トピック	50	
		総合評点	1	
		個別の観点に基づく評点	5	
	注目度	被閲覧数	1	
役に立った数		1		
アイテム付随特徴	メタデータ	公開年代	1	
		タイトルの長さ	1	
	内容	同監督の映画数	1	
		カテゴリ	15	
		映倫区分	1	
	人気度	レビュー被投稿数	1	
		「観たい映画」被登録数	1	
		「観た映画」被登録数	1	
	評判	総合評価	1	
	著者特徴	貢献度	総投稿数	1
			総文字数	1
			総閲覧数	1
			総獲得役立ち数	1
			平均文字数	1
		情報量	ネタバレレビュー率	1
平均閲覧数			1	
専門度		観点評価差平均	1	
		印象タグ数平均	1	
		平均獲得役立ち数	1	

表 3 用いた特徴ごとの分類器の分類性能

特徴	F_1
レビュー内在特徴 + アイテム付随特徴 + 著者特徴	0.771
レビュー内在特徴 + アイテム付随特徴	0.786
レビュー内在特徴のみ	0.568

判断し、対象の映画を観たいと思う度合いを 4 段階で答えた。

次に、分類器を用いて抽出したレビュー群を無作為な順で提示し、レビューをそれぞれ評価させた。評価は、「そのレビューを読むことでどれだけ映画を観たくなったか」を、4 段階で行った。

5.2 実験結果

レビューに内在する特徴と映画に付随する特徴と著者の特徴を用いた分類器、レビューに内在する特徴と映画に付随する特徴を用いた分類器、レビューに内在する特徴のみを用いた分類器について、分類結果と人手で付与したラベルの比較を行った。表 5 にそれぞれの分類結果に対する被験者評価の結果を示す。すべての分類器において、Student の t 検定による有意差検証の結果、分類器に「このレビューは人の視聴行動を誘発する」

表 4 特徴の重要度

仮説	注目要素	重要度
レビュー内在特徴	情報量	0.0560
	読みやすさ	0.0475
	内容	0.137
	注目度	0.0641
アイテム付随特徴	メタデータ	0.0388
	内容	0.117
	人気度	0.210
著者特徴	評判情報	0.0590
	貢献度	0.138
	情報量	0.0521
	専門度	0.0828
	合計	1.00

と分類されたレビューの方が、被験者が「そのレビューを読んで実際にその映画を見たくなった」と答える割合が有意に高いことが分かった。分類を行わず無作為に抽出したレビュー間では、有意差は見られなかった。

6 考察

はじめに、最も有効であると判定された、レビュー対象となる映画の特徴について考察する。分類器の学習には、レビューに内在する特徴と映画に付随する特徴と著者の特徴を用いた。表 4 に表される特徴の重要度から、映画に付随する特徴が視聴行動を誘発するか否かの推定に大きく寄与したことが分かった。このことから、人が映画を視聴するか否かは、映画に投稿されたレビューから受ける影響よりも、レビュー対象の映画による影響が大きかった。

注目要素のうち、映画の人気度が人の視聴意欲に最も大きく影響を与えることが分かった。人が映画の視聴を決定する要因として、そもそも映画が注目度の高い人気作品であるかが重要だと言える。注目のべき点として、映画の人気度は、映画の内容の要素や、映画の持つメタデータより大きい影響を持つことが挙げられる。このことから、人が映画を視聴するかどうかを決める要因として、どのような映画かという情報よりも、人気の映画かどうかという情報を重視していることが分かった。

また、映画に付随する特徴のうち、映画の評判情報がユーザの視聴行動に対して持つ影響は比較的小さいことが分かった。例えばある映画がレビューで酷評されている作品だったとしても、世間で注目されている作品ならば、人々はその映画を視聴するということが考えられる。

次に、レビュー内容の特徴について考察する。分類にレビューに内在する特徴のみを用いた分類器を用いて、視聴行動を誘発するレビューを特定することができた。このことから、レビューに内在する特徴のみを用いて、視聴行動を誘発するレビューを特定可能であることが分かった。

表 4 に表される特徴の重要度から、レビューに内在する特徴のうち、レビューの内容が人の視聴意欲に最も大きく影響を

表 5 分類器による「視聴を誘発する」,「視聴を誘発しない」と推定された各レビューに対する、被験者による「見たいと思った」度合いの評価結果 (*は $p < 0.05$, **は $p < 0.01$) .

映画	レビュー内容+映画+著者		レビュー内容+映画		レビュー内容		ランダム	
	上位 5 件	下位 5 件	上位 5 件	下位 5 件	上位 5 件	下位 5 件	上位 5 件	下位 5 件
キングダム	2.90*	1.70	3.50**	1.20	3.40*	2.50	2.90	2.70
トイ・ストーリー 4	2.90	2.70	1.40	2.40	1.70	1.60	1.40	1.60
名探偵コナン 紺青の拳	1.90	2.10	2.60*	1.50	2.60*	1.40	2.20	1.90
劇場版『ONE PIECE STAMPEDE』	2.70**	1.90	3.20**	1.70	3.10*	2.20	3.10	2.40
天気の子	2.50**	1.80	2.80	2.40	3.60**	2.10	3.50	3.60
アベンジャーズ/エンドゲーム	2.40	2.00	2.50*	1.50	2.40	2.20	2.20	2.90
アラジン	3.10**	1.40	3.20*	2.20	2.60	3.00	2.50	2.60
スター・ウォーズ/スカイウォーカーの夜明け	2.10	1.70	3.20**	1.70	2.60	2.50	2.10	2.00
アナと雪の女王 2	2.70*	1.60	3.00	2.60	3.30	3.00	2.90*	2.30
ライオン・キング	2.20	2.60	3.00	2.40	3.00*	2.00	3.20*	2.60
平均	2.54**	1.95	2.84**	1.96	2.86**	2.25	2.70	2.46

与えることが分かった。レビューの内容のうち、「物語の良さ」や「脚本の良さ」などの、個別の観点に基づく評点の特徴の重要度が最も高く、次に本文のトピックの特徴の重要度が高かった。このことから、ユーザはレビューを読む上で、何かしらの観点に基づいてレビューの内容から視聴行動を決めていることが分かった。また、レビューの情報量が視聴行動に与える影響は、レビューの内容や注目度の要素が与える影響よりも、小さいことが分かった。あるレビューが、他人に映画を勧める目的でレビューを書くとき、労力をかけ、長い文章などに代表される情報量の多いレビューを投稿するよりも、工夫した観点から映画を評価したレビューを投稿し、読み手に与える情報を考慮することが重要であることが分かった。

最期に、実際の分析上の問題点として、手法に改善の余地が残されていることも分かった。実験を行うにあたり、正解として使用したラベルの付け方が現実に即していなかった可能性がある。具体的には、多くの人が視聴する映画と、たかだか数人しか見ない映画が、同じようにデータセットに含まれた。この場合、映画によってレビューの持つ意思決定上の重要度が変わってくるが予想される。

ほかに、今回の実験ではログデータの使用上の制約から2か月間分に関してのみ分析した。そのため、たまたまその時期に無料公開された大作映画の特徴などが、行動を促すレビューの特徴に含まれた可能性がある。今後長期的なログを使って追加実験し、深く分析する必要がある。

7 まとめと今後の課題

本論文では、実サービスのアクセスログから、実際に読み手の行動を誘発したレビューの特徴を分析した。分析から得られた知見について、被験者実験によって、有用性を分析した。

今後の課題として、より大規模で長期的なログデータを使った分析、より多彩な特徴量を用いた分析、特徴量ごとのレビュー内容を定量的に分析する必要がある。

謝 辞

本研究はJSPS科研費18K18161(代表: 莊司慶行), 18H03243(代表: 田中克己)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] 藤村滋, 豊田正史, 喜連川優ほか. Webからの評判および評価表現抽出に関する一考察. 情報処理学会研究報告データベースシステム (DBS), Vol. 2004, No. 72 (2004-DBS-134), pp. 461–468, 2004.
- [2] Y. Zhou and S. Yang. Roles of review numerical and textual characteristics on review helpfulness across three different types of reviews. *IEEE Access*, Vol. 7, pp. 27769–27780, 2019.
- [3] Yue Pan and Jason Q Zhang. Born unequal: a study of the helpfulness of user-generated product reviews. *Journal of Retailing*, Vol. 87, No. 4, pp. 598–612, 2011.
- [4] Alton Chua and Snehasish Banerjee. Understanding review helpfulness as a function of reviewer reputation, review rating, and review depth. *Journal of the Association for Information Science and Technology*, Vol. 66, , 06 2014.
- [5] Susan M Mudambi and David Schuff. What makes a helpful review? a study of customer reviews on amazon. com. *MIS quarterly*, Vol. 34, No. 1, pp. 185–200, 2010.
- [6] Judith A. Chevalier and Dina Mayzlin. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, Vol. 43, No. 3, pp. 345–354, 2006.
- [7] Don Charlett, Ron Garland, Norman Marr, et al. How damaging is negative word of mouth. *Marketing Bulletin*, Vol. 6, No. 1, pp. 42–50, 1995.
- [8] Yuanyuan Hao, Qiang Ye, Yijun Li, and Zhuo Cheng. How does the valence of online consumer reviews matter in consumer decision making? differences between search goods and experience goods. pp. 1–10, 01 2010.
- [9] Hong Hong, Di Xu, G Alan Wang, and Weiguo Fan. Understanding the determinants of online review helpfulness: A meta-analytic investigation. *Decision Support Systems*, Vol. 102, pp. 1–11, 2017.
- [10] Pradeep Racherla and Wesley Friske. Perceived ‘usefulness’ of online consumer reviews: An exploratory investigation across three services categories. *Electronic Commerce Research and Applications*, Vol. 11, No. 6, pp. 548–559, 2012.

- [11] Zhiwei Liu and Sangwon Park. What makes a useful online review? implication for travel product websites. *Tourism Management*, Vol. 47, pp. 140–151, 2015.
- [12] Pei-Yu Chen, Samita Dhanasobhon, and Michael Smith. All reviews are not created equal: The disaggregate impact of reviews on sales on amazon. com. In *Carnegie Mellon University Working Paper*. 2008.
- [13] Yue Lu, Panayiotis Tsaparas, Alexandros Ntoulas, and Livia Polanyi. Exploiting social context for review quality prediction. In *Proceedings of the 19th International Conference on World Wide Web*, pp. 691–700, 2010.
- [14] Jiliang Tang, Huiji Gao, Xia Hu, and Huan Liu. Context-aware review helpfulness rating prediction. In *Proceedings of the 7th ACM Conference on Recommender Systems*, pp. 1–8, 2013.
- [15] L. Connors, S. M. Mudambi, and D. Schuff. Is it the review or the reviewer? a multi-method approach to determine the antecedents of online review helpfulness. In *2011 44th Hawaii International Conference on System Sciences*, pp. 1–10, Jan 2011.
- [16] Christy MK Cheung, Matthew KO Lee, and Neil Rabjohn. The impact of electronic word-of-mouth: The adoption of online opinions in online customer communities. *Internet research*, Vol. 18, No. 3, pp. 229–247, 2008.
- [17] Einar Bjering, Lars Jaakko Havro, and Øystein Moen. An empirical investigation of self-selection bias and factors influencing review helpfulness. *International Journal of Business and Management*, Vol. 10, No. 7, p. 16, 2015.
- [18] Chris Forman, Anindya Ghose, and Batia Wiesenfeld. Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information systems research*, Vol. 19, No. 3, pp. 291–313, 2008.
- [19] Phillip Nelson. Advertising as information. *Journal of political economy*, Vol. 82, No. 4, pp. 729–754, 1974.
- [20] Liqiang Huang, Chuan-Hoo Tan, Weiling Ke, and Kwok-Kee Wei. Comprehension and assessment of product reviews: A review-product congruity proposition. *Journal of Management Information Systems*, Vol. 30, No. 3, pp. 311–343, 2013.
- [21] Jing Luan, Zhong Yao, FuTao Zhao, and Hao Liu. Search product and experience product online reviews: An eye-tracking study on consumers’ review search behavior. *Computers in Human Behavior*, Vol. 65, pp. 420 – 430, 2016.
- [22] Julian John McAuley and Jure Leskovec. From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews. In *Proceedings of the 22nd international conference on World Wide Web*, pp. 897–908, 2013.