

# 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

櫛見 圭司<sup>†</sup> 北山 大輔<sup>†</sup>

<sup>†</sup> 工学院大学情報学部 〒163-8677 東京都新宿区西新宿 1-24-2

E-mail: [†j113035@ns.kogakuin.ac.jp](mailto:†j113035@ns.kogakuin.ac.jp), [kitayama@cc.kogakuin.ac.jp](mailto:kitayama@cc.kogakuin.ac.jp)

あらまし 近年、多くの映画が公開されており、その数は日々増大している。そのため多くの映画の中からユーザの好みの映画を見つけるのは難しい。一方で、映画に関する情報を公開する場としてレビューサイトがあげられる。レビューにはその映画のタイトルやあらすじなどの基本情報のほかに、視聴者による映画に対する感想や、映画を視聴して感じた印象、感情などが書かれて投稿されている。こうしたレビューにはその映画の特徴が多く含まれており、ユーザの好みの映画のレビューにはユーザの好む特徴が表現されている可能性がある。そこで、ユーザが選択した好みの映画のレビューを分析することで、ユーザの好みである特徴を含んだ映画を探そうと考えた。具体的には、映画のレビューに含まれる視聴者が感じた印象などの情報から、その映画に含まれる要素を特徴ベクトルとして算出する。その特徴ベクトルを他の映画の特徴ベクトルと比較し、同じ要素を多く含んでいる類似映画を推薦する手法を提案する。

キーワード 映画推薦, Doc2vec, ユーザレビュー

## 1. はじめに

近年、多くの映画が公開されておりその数は日々増大している。例として、Yahoo!映画<sup>(注1)</sup>に掲載されている映画数は59,439件にもなる。これらの映画の情報を知る方法として、映画.comやYahoo!映画などのレビュー投稿サイトがある。レビューサイトではユーザが映画を視聴した際の感想が公開されており、ユーザの視聴映画の決定に時に参考になっていると考えられる。レビューサイトを参考にする場合、予め視聴したい映画について調べる場合が多い。一方、ユーザが視聴する映画を絞り込めていない状態で好みに合った映画を見つけたい場合、レビューサイトを場当たり的に見るしかなく、好みに合う映画を探すのは困難である。この問題を解決するためには、ユーザが求める要素を抽出しその要素を含んでいる映画を推薦する必要がある。その方法の一つとして、ユーザの嗜好データやアイテムの特徴などを考慮し推薦を行うコンテンツベースフィルタリング[1]と呼ばれる手法がある。本研究もこの手法と同様に映画のレビューの内容に基づき、映画を推薦する手法である。

本研究では、映画推薦を行うためにユーザがどのような点を見て、視聴する映画を決めているのか調査した。被験者に映画5件の映画から1件選択してもらい、選択した映画から推薦されてきた映画3件に対して、見てみたいか見たくないか評価してもらった。その際、その評価にいたった理由を記述してもらった。その理由を著者らによって、「同様の特徴」、「異なる特徴」、「嫌いな特徴」、「興味ない特徴」、「俳優・監督」、「その他」の6種類に分類した。被験者は133人、推薦元となる映画は「インセプション」、「オデッセイ」、「ミスト」、「サマーウォーズ」、「ズートピア」の5件とした。推薦手法として、Yahoo!

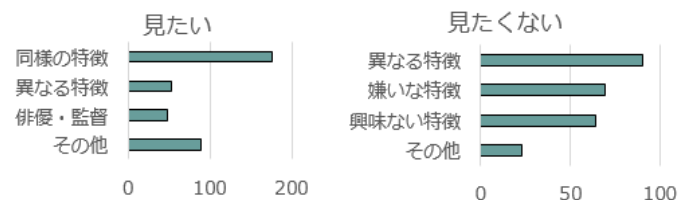


図1 映画の視聴決定理由

映画などのレビューサイトにある映画推薦機能によって推薦を行った。その結果、多くのユーザは好みの映画と同様の特徴を持った映画を見てみたいと回答した。また、見たくないと回答した多くは異なる特徴を持つ映画と回答した。結果を図1に示す。この結果より、映画の特徴を抽出しユーザの好む特徴を持つ映画を推薦することで、ユーザが求める映画を推薦できると考えた。そこで、レビューサイトに投稿された映画に対するレビューに注目した。

投稿されたレビューにはあらすじや出演者などの映画に関する情報だけでなく、投稿者が実際にその映画を視聴した時に感じた印象や感情、演出についての感想などが含まれている。これらの情報は投稿者が視聴した映画がどのような特徴を持つかを表す情報である。また、レビューには映画に関する多くの特徴の記述があり、一つのレビューに複数の映画の特徴が含まれている。この特徴を抽出することで、その映画がどのような要素を持っているのかを表現することができると考えられる。そこで、投稿されたレビューを用いて映画の特徴ベクトルを生成し、ユーザが選択した映画の特徴ベクトルと他の映画の特徴ベクトルの類似度を算出することで、同様の好みの要素を持つ映画を推薦できる手法を提案する。好みの要素を持った類似映画の提示を行うことで、短時間でユーザが目的としている映画

(注1) : <http://movies.yahoo.co.jp/>

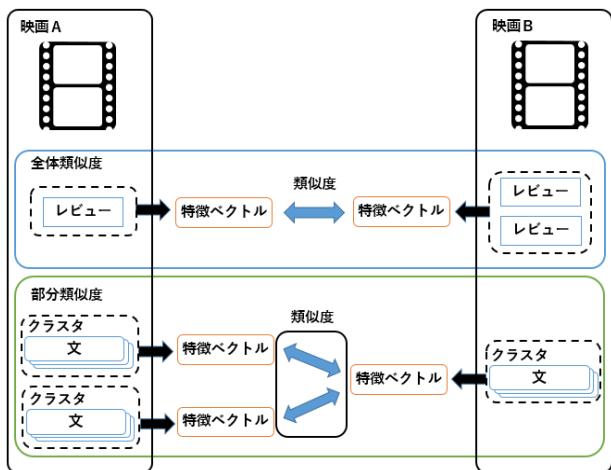


図2 提案手法の概要

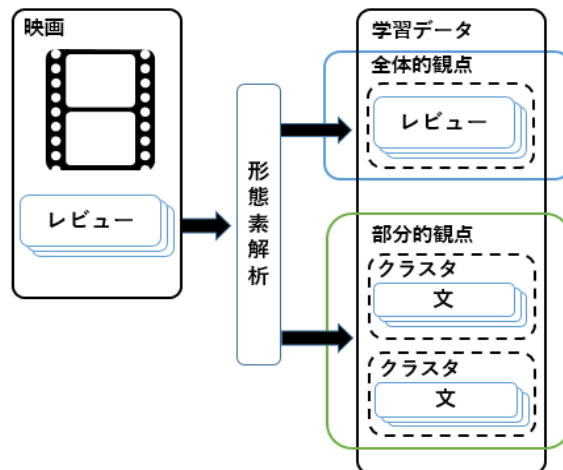


図3 レビューと学習データ

の候補を探し出せると考えた。提案手法の概要を図2に示す。  
 本稿の構成は以下の通りである。2節で関連研究を紹介する。3節では特徴ベクトルの類似性を用いた映画を推薦する提案手法や、提案手法に用いるデータや技術について説明をする。4節では実験を行いその結果について述べる。最後に5節でまとめと今後の課題について述べる。

## 2. 関連研究

映画推薦を行う研究として、映画に付けられているレビューの中の肯定興味語、否定興味語に着目し、ユーザが書いた映画のレビューから個人の嗜好などの判断を行い他者が書いた映画のレビューを推薦する手法[2]がある。また、ユーザの映画視聴履歴と映画評価からユーザの嗜好を抽出し、ユーザ特有の評価基準を考慮した手法[3]や、ユーザの映画の嗜好や興味を予測しそれらに沿った情報を推薦する手法[4]などレビューから映画に対する印象や評価などの特徴を抽出し推薦に利用する手法がある。このようにレビューやユーザの視聴履歴などから嗜好情報を抽出することで、ユーザの感性に合わせた映画を推薦することができる。

本研究では、関連研究[2][3][4]と同様に、レビューを分析し映画に対するユーザの感性を、定量的に表現した特徴ベクトルを用いることで推薦を行う。上記の関連研究では、単語や視聴履歴からユーザの嗜好や推薦する映画を決定しているが、本研究は映画の推薦にレビューを全体的、部分的に分析し特徴ベクトルとして算出する点で異なる。

また、レビューの特徴に関する研究として、特定の印象語群との共起関係に基づいて印象辞書を構築し、印象辞書から算出される記事の印象値と記事を読んだ人が感じる印象の強さとの対応関係に基づき、読み手が感じる印象の強さを定量的に予測する手法[5]や、レビューの文の位置や文脈構造の情報を利用することで、文中に含まれるあらすじ文の予測を行う手法[6]などがある。その他に、レビューやブログなどからユーザの体験を記述する際の特徴として現れる形態素の組み合わせを体験表現とし、その体験表現を手掛かりに体験情報を抽出する手法[7]

や、レビューの特徴を分類する方法として、クラスタリングを複数用いることで意味的特徴を抽出し、クラスタ内の意味一貫性を向上させることで、クラスタリング精度の向上を目指している手法[8]がある。これらの研究は、文章中の特徴の判定や検出精度の向上に貢献している。上記の関連研究[5][6][7][8]は、映画推薦に直接関わるものではないが、文章から映画に関する情報やユーザの映画の嗜好に関わる情報を抽出する手法として応用できると考えられる。

提案手法では、印象や興味語などの特定の単語だけではなくレビュー全体の単語、及び記述されている内容に基づいて文章ごとに分類した文を用いて映画の特徴ベクトルを生成し、ユーザの嗜好に合わせた映画を推薦する。

## 3. 特徴ベクトルと類似度

本章では、映画の特徴ベクトルの生成手法について説明する。また、全体的・部分的観点による特徴ベクトル間の類似度算出による類似映画の推薦手法について述べる。

### 3.1 映画の特徴ベクトルの生成

我々は映画のレビューによって、映画の特徴ベクトルを生成するために、Leら[9]によって提案された文書の特徴ベクトル獲得手法を用いる。この手法はMikolovらによって考案された単語特徴ベクトル獲得手法であるWord2vec[10]を文書に拡張したものである。Word2vecでは、「同じ文脈で出現する単語は類似した意味を持つ」というハリスの分布仮説[11]に基づき、「ある文書中である単語列が与えられた時、次に出現する単語を予測する」というタスクをニューラルネットワークに学習させることで、文脈や単語の語順を考慮した特徴ベクトルを獲得することができる。この単語学習結果を基に文書の特徴を学習するものである。この手法はgensim[12]<sup>(注2)</sup>パッケージ中にDoc2vecという名で実装されており、本研究でも以後Doc2vecと呼ぶ。このDoc2vecを用い映画のレビューから特徴ベクトルを生成する。本研究ではレビューを文書としてDoc2vecに

(注2) : <https://radimrehurek.com/gensim/>

用いる。Doc2vec は学習データとして単語単位に分ち書きされたテキストデータを必要とするため、形態素解析を用いてレビューを単語単位に分ち書きし、Doc2vec の学習データとして用いた。レビューと学習データの関係を図 3 に示す。

### 3.2 特徴ベクトルの類似度算出

Doc2vec により生成した各映画の特徴ベクトルには、映画に含まれる様々な要素の情報が含まれており、これらの特徴ベクトルと他の映画の特徴ベクトルとの類似度を算出することで、各映画に含まれる要素やレビュー投稿者たちが映画を視聴した際に着目した点や印象などが類似すると考えられる映画をユーザに推薦する。映画間の類似度には Cos 類似度を用いる。Cos 類似度はベクトル空間モデルにおいて、文書間を比較する際に用いられる類似度計算手法である。Cos 類似度はベクトル間の角度の近さを表現し、1 に近いほど類似しており 0 に近いほど類似していないことを示す。以下の式により、Cos 類似度を求める。  $x$  は求めたい要素を持っている映画の特徴ベクトル、  $y$  は比較を行う映画の特徴ベクトルである。

$$\cos(x, y) = \frac{x \cdot y}{\|x\| \|y\|} \quad (1)$$

この Cos 類似度を用い 2 つの映画の特徴ベクトル間の類似度を求めることで、特徴が類似すると考えられる映画の推薦を行う。また、本研究では映画間の類似性を求める際に、全体類似度と部分類似度の 2 つの類似度を用いる。

### 3.3 映画の全体的観点

映画のレビューには様々な特徴が含まれている。そうした特徴は映画ごとに含まれる傾向が異なると考えられる。例えば、アクション映画であれば、「カーレース」や「銃撃戦」などのストーリーに関する特徴や「ドキドキした」や「迫力がある」などのユーザの感じた特徴が多く含まれ、ドラマ映画であれば「青春」や「家族愛」などのストーリーに関する特徴や「感動した」や「幸せな気持ちになった」などの特徴が多く含まれる。このような映画ごとの特徴の類似性がある映画は、同様なジャンルやストーリーである映画と考えられる。こうした類似性のある映画は、単一映画の全レビューを用いることで求めることができると考えられる。単一映画の全レビューを用いることで、その映画に含まれる要素の類似性を総合的に比較することができ、各映画ごとの全体的な類似度を求められる。本研究では、ある映画の全レビューを用いて生成した特徴ベクトルから得られた類似度を、全体類似度として用いる。全体類似度を求める際の概要を図 4 に示す。

### 3.4 映画の部分的観点

レビューに含まれる情報は多いが、その全てが映画に関する情報とは限らない。映画の演出に関する情報や、映画を視聴した際の視聴者が感じた印象や感情などに関する情報が含まれている場合もあれば、視聴者がその映画を見に行った時の天気や視聴者自身の情報が書かれているなど、映画に関係のある情報ではない場合もある。この場合、ただ類似映画を求めるだけ

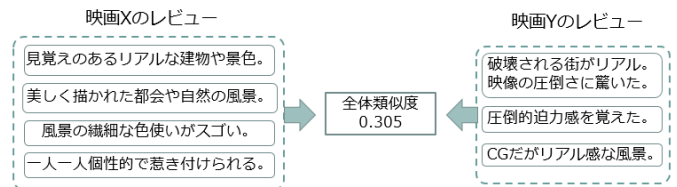


図 4 全体類似度の概要

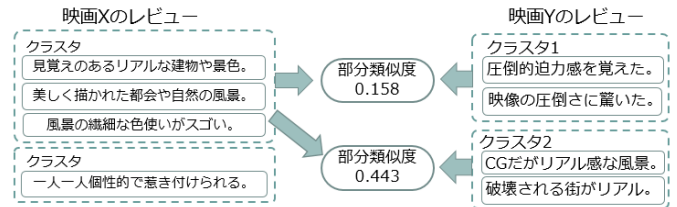


図 5 部分類似度の概要

では、映画に関係のない部分が類似した映画になってしまう可能性がある。また、単一映画の全てのレビューを用いて類似映画を求めた場合、比較対象の映画とどのような点で類似しているかを知ることは難しい。その他に、ある映画と同じような「カーレースでのアクション」が見たい場合など、特定の特徴を多く含む映画を求められないなどの問題がある。そこで、レビューを文ごとに分割し各文がその映画の何に対する文なのかを考慮することで、全体的に類似している映画とどのような部分が類似しているのかわかるようになり、また特定の特徴を重視した推薦が行えると考えられる。

映画の部分類似度は、単一映画のレビューを文ごとに分割し、文に記述されている内容や意味に基づいて分類することで求めることができると考えられる。一つのレビューには映画の演出や視聴者が感じた印象などに関する複数の要素について記述されており、これらの要素を基にレビューを分類することは難しい。そこで、レビューを文ごとに分割し同様の内容について記述されているように分類することで、各映画の同様の内容のものと比較することができ部分的な類似度を求めることができると考えた。

分割したレビューを分類する手法として、階層的クラスタリングを用いた。クラスタ間の距離測定方法に群平均法を用い、距離は Cos 距離を用いた。文章に分割したレビューごとの特徴ベクトルに対して階層的クラスタリングを行うことで、同様の内容について記述されるように分類しクラスタを生成した。本研究では、推薦元の映画のあるクラスタと他の映画のクラスタの特徴ベクトルから得られた類似度の中で最も値の大きいものを、部分類似度として用いる。部分類似度を求める際の概要を図 5 に示す。

### 3.5 映画間の類似度

2 つの映画間の最終的な類似度は全体類似度と部分類似度を合算することで求める。下記の式は 2 つの映画間の類似性スコアである。

表 1 ポリス・ストーリー レジェンド の全体的観点による推薦映画

映画名	類似度
ライジング・ドラゴン	0.314
ダブル・ミッション	0.310
九龍の眼	0.308
シビル・ウォー キャプテン・アメリカ	0.303

表 2 君の名は。 の全体的観点による推薦映画

映画名	類似度
言の葉の庭	0.400
秒速 5 センチメートル	0.385
ラッキー・ガール	0.333
雲のむこう、約束の場所	0.318

$$score(X, x) = \alpha \cos(X, Y) + (1 - \alpha) \max_{y \in Y_c} \cos(x, y) \quad (2)$$

推薦するために使用する映画を入力映画，入力映画をもとに推薦される映画を推薦映画とする。X, Y は入力映画，推薦映画の単一映画ごとの全レビューの特徴ベクトルで，x は入力映画のユーザが選択したレビュー文クラスタの特徴ベクトル，y は推薦映画のあるレビュー文クラスタの特徴ベクトルであり，Y<sub>c</sub> はその集合である。これらの Cos 類似度により全体類似度，部分類似度を算出する。α は全体類似度，部分類似度をどの程度重視するのかの重みとする。

#### 4. 予備実験：推薦手法の特性の確認

##### 4.1 実験方法

提案手法によって提示される映画にはどのような傾向・特性があるのかを確認するために映画.com から収集したレビューに提案手法を適用した。推薦するために使用する映画を入力映画，入力映画をもとに推薦される映画を推薦映画とする。入力映画は任意の映画の一つを選び，入力映画と類似していると判断された推薦映画を提示する。実験で用いるデータは映画.com から取得したレビューを使用する。映画タイトルは 11,358 件，総レビュー数は 154,454 件である。入力映画としては「ポリス・ストーリー レジェンド」「君の名は。」の 2 件を用いた。特徴ベクトルの算出は gensim を使用し，形態素解析した各映画のレビューを学習データとして用いた。レビューの形態素解析は MeCab [13] を利用し学習データを作成した。クラスタリングの際の Cos 距離の閾値は 0.5 としてレビューを分類した。

##### 4.2 全体類似度 (α = 1.0) による推薦結果

表 1, 2 はレビュー全体の観点の特徴ベクトルに基づいた，各入力映画から得られた推薦映画の映画名と類似度である。表 1 は，アクション映画である「ポリス・ストーリー レジェンド」を入力映画として得られた推薦映画である。類似度の上位のものは全てアクション映画でありジャンルなどが一致している映画が多く，映画の基本情報では類似性が見られた。映画の内容を確認した所，上位 3 つの映画は出演俳優がジャッキー・チェンであり，入力映画である「ポリス・ストーリー レジェンド」

表 3 ポリス・ストーリー レジェンド の部分的観点による推薦映画

映画名	類似度
プロジェクト A	0.409
パーフェクト・ガイ	0.325
ブルドッグ	0.323
パッチ・アダムス トゥルー・ストーリー	0.321

表 4 君の名は。 の部分的観点による推薦映画

映画名	類似度
小さいうち	0.495
ニコラス・ケイジのウェザーマン	0.438
恋は嵐のように	0.433
僕等がいた 後篇	0.421

表 5 ポリス・ストーリー レジェンド のクラスタの文章

物語の核心を後半まで引っ張り続けるストーリーは想像以上にスリリングに描かれる。
最後の最後まで正義一直線のジャッキーに感動。
確かに、作り方よっては心理サスペンスにもなりそうなストーリーだ。

表 6 君の名は。 のクラスタの文章

しかし、恋ってやっぱり良いなどとも思わせてくれる作品でした。
終わり方もすごく良かった!
笑い、泣き、王道の展開、でも新海監督じゃなきゃ描けないファンタジーと現代ものが見事なまでのミックス。

の出演俳優もジャッキー・チェンであった。そのためこれら 3 つの映画は出演俳優と，同じジャンルであるアクション映画という点で高い類似性を示している。また，別の同ジャンル作品に出演した俳優の演技などと比較した意見や感想と，アクションやカメラアングルなどの演出に関する意見や感想の両方，あるいは片方を含んでいるレビューが多く見られた。

表 2 では，アニメ映画である「君の名は。」を入力映画として得られた推薦映画である。推薦映画は入力映画と同じジャンルであるアニメ映画が多く見られた。上位 2 件と下位 1 件はアニメ映画であるが，これらと入力映画の共通点として監督が挙げられる。これらの作品を手がけた監督が「新海誠」であり，監督に関する記述がレビューに数多く見られた。一方，「ラッキー・ガール」はアニメ映画ではなく，また監督が「新海誠」ではない。この映画と入力映画の類似性を調べたところ，「ラッキー・ガール」は男女の運勢が入れ替わる（不幸な男性と幸運な女性）が，ふとしたことでそれぞれの運勢が入れ替わるというストーリー）であり，「君の名は。」の主要な要素である「男女の入れ替わり」という点で類似していることが分かった。

##### 4.3 部分類似度 (α = 0) による推薦結果

表 3, 4 は部分的観点の特徴ベクトルに基づいた，各入力映画から得られた推薦映画の映画名と類似度である。表 3 は表 5 の文章が含まれるクラスタの特徴ベクトルの類似度により推薦された映画である。表 5 のクラスタは，緊迫感のあるアクション

ンシーンを見てハラハラすることやそのシーンを見て感動したなどの意味を持つ文章が見られた。そのため、アクションやサスペンスなどの要素を含む映画が推薦されている。推薦映画の一つである「パーフェクト・ガイ」の類似性の高いクラスタには「冷静にみると穴は結構あるけれど、ハラハラドキドキ楽しめた。」という文章があり、緊迫するなどの要素を持つ文章が分類されている。

表4は表6の文章が含まれるクラスタの特徴ベクトルの類似度により推薦された映画である。表6のクラスタは恋愛に対する意見や泣けた、感動したなどの文章が多く見られた。そのため、恋愛映画や家族愛などの要素を持つ映画が中心に推薦された。推薦映画の一つである「ニコラス・ケイジのウェザーマン」のクラスタには「この映画の観終わった後の切ない感じが僕は好きです」という文章があり、泣けるなどの要素を持つ文章が見られた。

#### 4.4 考 察

レビュー全体的観点に基づく類似性による映画推薦では、ある程度類似点を持った映画を推薦することができる結果が得られた。レビュー全体を用いた推薦の場合、多くのケースで登場人物や監督といった特定の人物名などの固有名詞が類似性に強く現れることが分かった。例えば出演俳優である「ジャッキー・チェン」や、監督である「新海誠」など、その作品に関わる有名な人物が各映画に存在している。これらの人物名はレビュー中に多く見られるため、これらの人物名が含まれる映画であれば、全体的に高い類似性が得られると考えられる。

また、表2の「ラッキー・ガール」のように、その作品を表す主要な要素（今回の場合「男女の入れ替わり」）などでも、高い類似性が見られた。これらのケースから、全体的な観点に基づく類似性では、その映画の主要な要素、あるいは有名な人物名などの固有名詞などが強く影響されていることがわかる。そのため、全体的な観点の類似性に基づく映画推薦では頻出する要素が強く反映され、個々が感じた印象や感情、またはカメラアングルやエフェクトなどの映画の演出に関する要素が反映されにくいという問題が生じる。

レビューの部分的観点に基づく類似性による映画推薦では、入力映画のクラスタの特徴を持った類似映画を推薦することができる結果が得られた。レビューの部分的な文章を用いた推薦の場合、全体的観点の場合とは異なり映画のジャンルや出演俳優などにあまり影響を受けず、クラスタの持つ特徴により全く異なるジャンルの映画が推薦されることが分かった。しかし、クラスタ内の文章の意味の一貫性が取れていないクラスタもあり、推薦を行うためには意味が一貫しているクラスタを探す必要がある。

### 5. 評価実験：推薦映画の特徴保有精度の評価

#### 5.1 実験方法

評価実験では、提案手法による推薦映画に入力映画が持つ特徴がどの程度含まれているのか検証することを目的とする。 $\alpha$

表7 各手法の特徴保有精度

手法	特徴保有割合	上位特徴保有割合
従来手法1 (Yahoo!映画)	0.506	0.318
従来手法2 (amazon)	0.663	0.355
提案手法 ( $\alpha = 0.3$ )	0.801	0.478
提案手法 ( $\alpha = 0.5$ )	0.747	0.437
提案手法 ( $\alpha = 0.7$ )	0.765	0.450

の値を変更することで、部分的特徴の保有精度がどのように変化するのか、また部分的特徴をどの程度考慮するのが適切なかを考察する。また、Yahoo!映画の映画推薦機能による手法と、amazonの商品推薦機能による手法の2つの映画推薦システムによる推薦映画の特徴保有精度と比較して評価する。

本実験では83人の被験者に、入力映画と各手法によって推薦された映画を2件ずつ、計10件を提示し推薦に用いた映画と同様の特徴を含んでいると思う映画を選んでもらった。また同時に入力映画の特徴がより一致する順番に映画を順位付けを行った。このとき、どの手法により推薦された映画かは被験者にわからないように提示した。被験者には入力映画、推薦映画に関する情報として「映画.com」の各映画のページを提示し特徴を保有しているか検証した。それ以外のパラメータについては、4章で用いたものと同様である。

#### 5.2 特徴保有精度に関する結果と考察

各手法により推薦された映画の特徴保有割合を示したのが表7である。表の特徴保有割合は、各手法によって推薦された映画に入力映画のレビュー文の要素が含まれていると評価された映画の割合を表す。また、上位特徴保有割合は各手法で推薦された映画の中で、入力映画の特徴と一致する映画を順位付けた際に、上位3件に選ばれた割合を表す。提案手法による特徴保有割合は従来手法と比べ、2割から3割ほど高い結果が得られた。また、上位特徴保有割合に関しても提案手法の方が高い結果となった。この結果から、提案手法では推薦元の映画が持つ特徴を考慮した推薦が行えていることがわかる。従来手法では映画の中身やユーザのレビューなどが考慮されず、ユーザがどのような点でその映画を好んでいるのかといった情報を利用してきていない。一方で、提案手法では映画レビューの中身を表現していると考えられる。

また、 $\alpha = 0.3$ が最も特徴保有精度が高い結果となった。これは、映画の持つ部分的特徴をうまく抽出できており、部分的特徴を多く含む映画の推薦を行うことができたためと考えられる。一方で、 $\alpha = 0.7$ が次に特徴保有精度が高い結果となった。これは、全体的特徴の中でも特に強い特徴が、推薦に用いられた特徴と一致、もしくは類似するものであるためだと考えられる。例えば、本実験で入力映画にアニメ作品である「時をかける少女」を用いた際、最も特徴保有割合が高いのは $\alpha = 0.7$ の推薦映画である「君の名は。」であった。この推薦に用いたクラスタには「若い彼と再開するときですごく気になる部分ですよね。」や「だれしもが過去に経験、あるいは今その真っ只中にある『青春』がぎゅっと詰まっています。」などの、両映画

の全体的特徴が多く含まれると考えられる特徴をもつクラスタを部分的特徴として用いていた。そのため、 $\alpha$  が 0.7 と高い場合でも、部分的特徴を保有していると評価された結果となったと考えられる。これらのことから、その映画の代表的な特徴を重視した推薦する場合  $\alpha = 0.7$ 、それ以外のより細かい特徴を重視した推薦する場合  $\alpha = 0.3$  にすることで、よりユーザの求める特徴を含んだ映画を推薦できると考えられる。

## 6. おわりに

本稿では、レビューにはあらずじや出演者などの映画に関する基本情報の他に、映画を視聴した際に感じた印象や感情、及びカメラアングルやエフェクトなどの映画の演出などに対する感想などが含まれていると考え、それらの特徴の類似性から類似映画を推薦する手法を提案した。提案手法では、いくつかの高い類似度を持つ映画を提示することができた。

今後の課題として新たな文章の分類手法の模索があげられる。今回行った階層的クラスタリング手法により分類された文章は、別々の意味を持った文章が混ざったクラスタが多く見受けられた。そのため現在の階層的クラスタリングによる分類では、効果的な文章の意味的分類を行うことができなかった。その原因として、部分的な観点による特徴ベクトルでは、情報量が少ないことが考えられる。レビューを文章に分割した特徴ベクトルでは文章量が少ないため、映画の要素などを表す単語の意味を推測する情報がとても少なく、うまく特徴ベクトルとして抽出できずに正しい意味的一貫性を持ったクラスタの生成ができなかったと考えられる。そのため部分的な観点による映画間の類似性の精度の向上について、検討していく必要がある。

その他に、有効性の検証で提案手法と比較した手法がどのようなことを行っているのか明らかになっていないため、提案手法の有効性に疑問がある。今後、有効性が明らかな他の推薦手法との比較を行い提案手法の有効性を検証を行っていく必要がある。

## 謝 辞

本研究の一部は、平成 28 年度科研費若手研究 (B)(課題番号: 15K16091) によるものです。ここに記して謝意を表すものとします。

## 文 献

- [1] 敏弘神島. 推薦システムのアルゴリズム (2). 人工知能学会誌, Vol. 23, No. 1, pp. 89–103, 2008.
- [2] 貴宏林, 理紀夫尾内. Web 上のレビューを利用した映画推薦システム. 人工知能学会論文誌, Vol. 30, No. 1, pp. 102–111, 2015.
- [3] 村中清史, 横井健. 映画のメタデータに対するユーザ特有の重要度を用いた映画推薦手法の提案. 第 76 回全国大会講演論文集, Vol. 2014, No. 1, pp. 619–620, 2014.
- [4] 英雄廣瀬. Netflix データベースにおける映画推薦アルゴリズムと嗜好予測の精度. 日本計算機統計学会大会論文集, No. 23, pp. 77–80, 2009.
- [5] 熊本忠彦, 河合由起子, 張建偉. ユーザ印象評価データの観察と分析に基づくテキスト印象マイニング手法の設計. 研究報告自然

- 言語処理 (NL), Vol. 2012, No. 4, pp. 1–10, 2012.
- [6] 秀成岩井, 嘉徳土方, 正吾西田. レビューの文脈一貫性を用いたあらずじ文判定手法. 情報処理学会論文誌データベース (TOD), Vol. 7, No. 2, pp. 11–23, 2014.
- [7] 佳代池田, 勝義田邊, 英範奥田, 雅博奥. Blog からの体験情報抽出. 情報処理学会論文誌, Vol. 49, No. 2, pp. 838–847, 2008.
- [8] 進也佐藤, 公海高橋, 真人松尾. 特徴抽出を目的とした文書クラスタからの一貫性阻害要素除去. 情報処理学会論文誌データベース (TOD), Vol. 6, No. 3, pp. 1–12, jun 2013.
- [9] Quoc V. Le and Tomas Mikolov. Distributed representations of sentences and documents. In *Proceedings of the 31th International Conference on Machine Learning, ICML 2014*, pp. 1188–1196, 2014.
- [10] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pp. 3111–3119. Curran Associates, Inc., 2013.
- [11] Zellig S. Harris. Distributional structure. *WORD*, Vol. 10, No. 2-3, pp. 146–162, 1954.
- [12] Radim Řehůřek and Petr Sojka. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pp. 45–50, Valletta, Malta, May 2010. ELRA. <http://is.muni.cz/publication/884893/en>.
- [13] Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. Applying conditional random fields to japanese morphological analysis. In *Proceedings of EMNLP 2004*, pp. 230–237, 2004.