定番映画の印象ベクトルに基づく映画推薦手法の提案

鈴木　悠太†　　佐藤　哲司††

† 筑波大学情報学群知識情報・図書館学類　〒 305–8550 茨城県つくば市春日 1-2
†† 筑波大学図書館情報メディア系　〒 305–8550 茨城県つくば市春日 1-2
E-mail: †{ysuzuki,satoh}@ce.slis.tsukuba.ac.jp

あらまし　現在、膨大な数の映画タイトルに対して、ジャンルや出演者名などを入力として検索する映画検索サイトが普及しているが、様々な条件を組み合わせて所望のタイトルを検索するのは難しく, 特に, その映画の存在も知らないようなタイトルを発見することは困難であった. 映画は単なる娯楽の対象となるだけでなく、ユーザの様々な状況に応じて笑いや感動などを与えてくれる. 本論文では、そういった映画に関する感想や意見に含まれる印象情報と、多くのユーザが推薦する定番映画のタイトルとを用いて, 定番映画を与えることで未知あるいは所望の映画タイトルを検索する映画推薦手法を提案する. 実際に定番映画を用いて映画推薦を行った結果, 類似度の高い映画を示すことができたのでその結果を示す.

キーワード　映画推薦, 映画レビュー

1. はじめに

現在，Yahoo!映画, 映画.com などの映画検索サイトが多く存在している. この様な検索サイトの多くは, 作品タイトル名やキャスト名, 監督名など映画の属性情報を用いて検索することができる. しかし, 映画を検索する際に, その映画に関する属性情報を知らなければならず, また多くの映画の中から所望の映画を選択するには, 様々な属性情報を組み合わせて検索しなければならなかった.

映画は単なる娯楽の対象となるだけでなく, ユーザの様々な状況に応じて笑いや感動などを与えてくれるため, 映画を検索する際に属性情報を用いず, 映画鑑賞時に得られる印象情報を用いて検索したいという要求もある. このようなユーザの要求を満たす研究として, ユーザの状況や気分などの印象情報を入力とするものが知られており, 印象情報は映画を探す特徴の 1 つとして有効であることが示されている. 例えば「楽しい気分になれる映画」や「絶対に泣ける映画」に含まれる「楽しい」や「泣ける」といった表現が印象情報に相当する. しかし映画における印象には個人差があり, ある映画を鑑賞したユーザの印象表現が他のユーザの印象表現と必ずしも一致するとは言えない. そのため映画の印象を入力とした映画検索システムでは, ユーザが思っていた印象を持つ映画を検索することが難しかった. 本論文では, 映画を鑑賞したユーザの事を経験ユーザ, 映画を鑑賞していないユーザを未経験ユーザと定義する.

本研究では, 属性情報や印象情報を入力とせず, 映画鑑賞時に得られる複数の印象から映画を検索できる手法を提案する. 提案法では, これまでの属性による検索の困難さを解決すること, さらには印象における個人差をなくすことを目的とする. そこで, 多くの経験ユーザから推薦され, 複数の印象が含まれている定番映画を例示し, これを検索の入力とする. そして定番映画に含まれる印象と類似した印象を持つ映画を推薦する. この一連を実現するために, NAVER まとめを用いる.

NAVER まとめ (注1) とは, ユーザがあらゆる情報をまとめて自由に組み合わせ, ひとつのページに保存・紹介できるサービスである. NAVER まとめの中で映画に関する情報がまとめられている映画トピックに着目する. 映画トピックでは映画に関する記事タイトルが多数存在し, その中に複数の映画が整理・体系化されている. 例として「笑える映画 10 選」が NAVER まとめ記事の記事タイトルに当たり, その中にユーザが「笑える」と感じた映画が 10 本紹介されている.

NAVER まとめの問題点として, 映画と関連度の低い記事の出現や, 印象を含むタイトルの不規則な羅列が挙げられ, 映画を探し出したい未経験ユーザにとっては負担となってしまう. そこで本研究では NAVER まとめに掲載されている映画に関する記事をデータベースとして扱い, 未経験ユーザでも使いやすい映画の検索と推薦を目指す.

第 2 章では関連研究を紹介し, 本研究の位置づけを明確にする. 第 3 章では定番映画の推薦手法を提案する. はじめに記事タイトルと映画の取得方法, 記事タイトルから印象表現の抽出方法を述べ, 感情の付与, また取得した映画名の修正方法について述べる. 第 4 章では印象ベクトルに基づく映画の推薦とその考察を行い, 第 5 章では本論文のまとめを述べる.

2. 関連研究

2.1 映画推薦システムについての研究

映画推薦システムについての研究は盛んになされている. 小野らはユーザの特性に関する情報とユーザの嗜好である気分や場所などの複雑な依存関係を表現したベイジアンネットモデルに基づく映画推薦システム [1] を提案している. このシステムでは映画を検索する際に, 状況と気分を入力し, それに合った映画が推薦される. ユーザの嗜好である気分を映画検索に用いる点で, 本研究と類似しているが, ユーザの気分や場所を入力する点

(注1)：NAVER まとめ http://matome.naver.jp/

で, 属性情報や印象情報を入力としない映画検索を目指す本研究とは異なっている.

また Miller ら [2] は携帯端末での映画推薦システムを研究しており, ユーザが映画を買うあるいは借りる場合, 映画館で映画を見る場合の 2 つの状況下で推薦を行い, 映画推薦の有用性を示している.

上松ら [3] は Linked Data に HITS アルゴリズムを適用したユーザの興味に沿った映画推薦システムを開発している. このシステムでは俳優, 脚本家, 監督の 3 つのうちいずれかを重視するユーザの趣向に合わせた映画推薦を行っている. 例えば「映画は俳優で決め, 監督・脚本家はほとんど気にしない」や「映画は監督・脚本家で決め, 俳優はほとんど気にしない」など, ユーザの優先したい趣向で映画推薦を行っている. 俳優や監督名などの属性情報は一切用いず, 複数ユーザの印象を重視して映画推薦を目指す本研究とは異なる.

2.2 感情モデルの比較

本研究で必要となる映画に関する印象表現は感情表現と似ているため, 感情を抽出する研究とも関連する. 感情表現を抽出する研究は多数行われており, その中で複数の感情表現を表す多次元の感情モデルが提案されている. 代表的なものとして, Plutchik [4] のモデルがある.「Joy ⇔ Sadness」「Acceptance ⇔ Disgust」「Fear ⇔ Anger」「Surprise ⇔ Anticipation」の 8 感情 4 次元のベクトルで表されており, さらに 8 つの基本感情から 2 つの感情を組み合わせることで, より細かな感情を表すことができるとするモデルである.

中村 [5] の感情モデルでは, 様々な日本語の表現を「喜, 怒, 哀, 怖, 恥, 好, 厭, 昂, 安, 驚」の 10 個の感情軸に当てはめ, 辞典を生成している.

このように様々な感情モデルが存在するが, 本研究では中村の感情を用いる. その理由として, 本研究で扱う印象表現が日本語であること, 中村の辞典に含まれる表現の多くが映画に関する印象表現に当てはめることができることが挙げられる.

2.3 印象抽出に関する研究

前節で述べたように感情モデルを用いて印象を抽出する研究は多数存在し, テキストから印象表現を抽出する研究として熊本ら [6] の研究がある. この研究では Plutchik の感情モデルを用いて印象辞書を自動構築し, 印象の抽出を行っている. その際にまずテキストを形態素解析し単語に分解する. そしてテキストに含まれる印象語の共起頻度を求める事で, 印象語をそれぞれの感情に自動的に分類している. 印象辞書を構築する際に印象語を登録していく作業が必要だが, 熊本らの印象語に感情を自動的に分類する手法を用いることで, 印象辞書に印象語を登録する時間を大幅に短縮することができる.

中村の辞典には映画の印象を表せる印象表現が既に多く存在しており, 中村の辞典を参照することで, 印象語の登録時間を大幅に削減できる. 本研究では中村の辞典に含まれる印象語を参照しながら, 映画特有の印象表現で不足しているものを印象辞書に手作業で追加する.

2.4 本研究の位置づけ

本研究では属性情報や印象情報を入力とせずに, 多くのユーザから推薦されている代表的な映画を例示することで, それと類似した新たな映画を推薦する. そのため映画を例示するという点で, 気分や場所を入力する小野らのシステムや監督名, 俳優などの確定情報を入力する上松らのシステムとは異なっている. 印象抽出においても, 印象語に感情を自動的に分類して印象辞書を構築する熊本らに対し, 中村の辞典に既にある日本語表現を印象語とし, 手作業で印象辞書を構築する点で異なる. 本研究では, 個人差があり言葉で表す事の難しい印象や, 監督名やキャスト名などの属性情報を入力とせず, 複数の印象を含む代表的な映画を例示することで, それと類似した映画を推薦するという手法に新規性があると考える.

3. 定番映画の取得法の提案

本章では, 定番映画の取得法を大きく 4 節に分けて述べる. まず 3.1 節で NAVER まとめから記事タイトルと映画を取得する方法について述べ, 3.2 節では取得した映画名の修正方法を述べる. 3.3 節では印象を抽出するに当たって, 必要な印象辞書の構築と印象語の追加について述べ, 3.4 節では生成した印象辞書を用いて, 記事タイトルと印象語を照らし合わせ, 映画に感情を付与する方法を述べる. 多くのユーザから推薦された映画のうち上位 100 件を定番映画とし, 上位 20 件を表 6 に示す.

3.1 記事と映画のタイトル取得

本研究では, 経験ユーザが生成した NAVER まとめ記事タイトルに含まれる印象表現を利用する. NAVER まとめ記事の構成として, ユーザが編集した記事タイトル名があり, そこに様々な情報がまとめられている. その概要図を示す.

図 1 NAVER まとめ記事の構成

NAVER まとめから記事タイトルを取得するため, NAVER まとめで映画に関する記事をまとめた「映画トピック」に属する記事から取得を行った. その結果, 記事タイトルを 6,564 件, それに付随する映画を 52,500 件取得できた. 取得した全データは映画データベースに格納し, そこに記事タイトルを格納する「title」カラムと映画を格納する「movie」カラムを作成した.

3.2 取得データにおける表記揺れの統一

本研究は多くの経験ユーザから推薦され, 多くの印象が含ま

れている定番映画を利用するため, 定番映画を定める必要がある. そのために, 映画データベースに格納された映画名と映画の数を調べることで定番映画を決定していく. しかし NAVER まとめから取得したデータは, ユーザの手によってまとめられているため, 表記が統一されていない. 例えば映画名の前に「○」や「▼」などの記号がついているもの, 映画名の前に「第 1 位」の文字がついているもの, そもそも間違った映画名でまとめられているものなどが多数存在しており, 同じ映画のはずが違う映画として認識されてしまうという問題がある. そのため NAVER まとめから取得してきたデータで定番映画を決定する事は難しく, 表記の揺れを修正する必要がある.

そこで, 取得データにおける表記揺れの統一を行うため, 映画名を修正する. 1 つ 1 つの映画名を修正することは膨大な時間を要するため, 表記揺れの中でもある程度規則的な表記揺れを MYSQL を用いて修正した. 用いた SQL 文の一例を表 1 に示す.

(1) の SQL 文は映画名が間違っているものを正しい映画名に直す SQL 文である. 映画データベースの中には映画名の表記が統一されていないものが多数存在している. 例えば「アイ・アム・サム」と「アイアムサム」では違う映画として認識されてしまう. このような映画名の違いは 1 つの映画名に統一する必要があり, 間違った映画名を修正する場合にはこの文を用いた. また正しい映画名は映画検索サイトである Yahoo!映画 (注2) のデータベースから取得し, 修正を行った.

(2) の SQL 文は特別な記号を取り除く SQL 文である. この例文では括弧でくくられている映画に対し, 括弧を取り除き映画名だけを残している. その他にも「♡ ラブ・アクチュアリー」や「▼メリーに首ったけ」のような, 映画名の前にハートや三角記号などの特別記号がついているもの, 「インセプション (2010)」のような, 映画名の後ろに公開された年数がついているものが多数存在する. 上述したような特殊記号の削除は, この文を用いて行った.

表 1　表記揺れの修正を行うときに用いた SQL 文の一例

```
(1) UPDATE eiga set movie = ' 正しい映画名'
    where movie like ' 間違った映画名';
(2) UPDATE eiga set movie =
    substr(movie,2,(char_length(movie)-2))
    where movie like '【%】'
```

3.3　印象表現辞書の構築

記事タイトルに含まれる印象表現を抽出するには, 6,564 件の NAVER まとめ記事タイトルの中から, 映画の印象を表している単語を抽出する必要がある. しかし映画の印象を表す単語は多数存在し, ユーザによって表現の仕方が異なるという問題がある. これを解決するため, 多くの印象語を用いて記事タイトルから印象表現を抽出する必要がある. そこで中村 [5] の『感情表現辞典』を用いて印象語の取得を行った. 中村の辞典では「喜, 怒, 哀, 怖, 恥, 好, 厭, 昂, 安, 驚」の 10 個の感情軸を用いて, 様々な日本語の表現を 10 個の感情にグルーピングしている. 例えば「喜」という感情カテゴリには「楽しい」「笑う」などの表現が載っており, 本研究ではこの「楽しい」「笑う」を印象語として用いて, 多くの印象表現を網羅する. そこで中村の辞典に記載されている印象語を登録していくことで, 印象辞書を構築した. 印象語の感情はそのまま中村の 10 個の感情軸を用いた.

中村の 10 個の感情軸を用いるため, 記事名や映画名を格納している映画データベースに感情カラムを生成した.

3.4　印象語の追加

あらかじめ記載されている中村の辞典の印象語だけでは, NAVER まとめ記事タイトルで表現されている映画特有の表現に対応することが出来ない. 例えば, 映画のカテゴリを示す「コメディ」や「ホラー」などの表現や, 映画の作風を表す「グロい」や「胸キュン」などの表現がこれに当たる. そこで NAVER まとめ記事タイトルで表現されている言葉にも対応できるようにするため, 中村の辞典に含まれている表現の他に, NAVER まとめ記事タイトルによく現れる以下の表現に対し感情を付与した.

(1) 中村の辞典の表現と類似している表現
(2) 明らかに感情が推測できる表現

(1) については, 例えば「胸キュン」という表現は, 中村の「好」カテゴリにある「恋心」と類似した表現であるため, 「好」の感情を付与した. (2) は, 例えば「ホラー」という単語は明らかに恐怖を意味する単語であるため, 中村の「怖」の感情を付与した.

また NAVER まとめ記事タイトルは様々な印象語の活用形やカタカナ, ひらがなの表記の違いが存在するため, 様々なパターンの印象語を網羅する必要がある. 例えば「驚」の感情をもつ印象語の「どきどき」は, カタカナの「ドキドキ」も印象辞書に登録する. また「喜」の感情をもつ「面白い」は「面白かった」や「おもしろすぎる」などの活用形が多く存在するため, 「おもしろ」や「面白」を辞書に登録している. 中村の辞典で映画の印象を表すことができる印象語と感情一覧を表 2 に示し, それに本節で追加した印象語とそれに対する感情一覧を表 3 に示す. 印象語を追加していった結果, 97 個の印象語を印象辞書に登録した.

(注2)：Yahoo!映画 http://movies.yahoo.co.jp/

表 2　中村の辞典の感情と印象語

感情	印象語群
喜	愉快, 楽しい, 明るい, 爽快, すっきり
怒	面白くない, 殺意, 不愉快, 気持ちが悪くなる
哀	悲しい, 泣く, しんみり, 虚しい
怖	怖い, 不安, 恐怖, どきどき, はらはら
恥	照れくさい
好	愛, 友情, 恋心, 恋
厭	鬱, 後味が悪い, 不愉快, 暗い, 苦悩, 困惑
昂	緊迫, 緊張, 揺れる, 躍り上がる, 感動, 興奮
安	和む, 落ち着く、安らぐ
驚	驚く, 仰天, ショック, 驚愕

表 3　印象表現辞書に追加した印象語の例

感情	印象語群
喜	笑, 面白, コメディ, 元気, おもしろ
哀	沁みる, 涙, 泣
怖	ホラー, ドキドキ, ハラハラ
好	恋人, 結婚したくなる, 胸キュン
厭	グロ, バイオレンス, キツイ, 衝撃
昂	人生を変える, 考えさせられる
安	美しい, ほのぼの, 癒
驚	どんでん返し, 結末が分からない

3.5　感情の付与

感情付与の方法として, 感情が付与された印象語が記事タイトルに含まれていた場合, 対応する感情のカラムに 1 を与え, それ以外を 0 とする. 例えば NAVER まとめ記事タイトルが「笑える恋愛映画 5 選」の場合,「喜」の感情をもった印象語である「笑える」と「好」の感情をもった印象語である「恋愛」がタイトルに存在しているため,「笑える恋愛映画 5 選」に含まれる映画 1 つ 1 つに対し,「喜」と「好」のカラムに 1 を与え, それ以外を 0 とする. また本論文では, このそれぞれの感情カラムに 1 もしくは 0 を付与した値を「印象ベクトル」と呼び, 第 4 章ではこの値を用いて映画推薦を行っていく. 映画データベースと各感情カラムに付与した印象ベクトルの一例を表 5 に示す. このように処理を行うことで, 52,500 件のうち 13,088 件の映画に感情を付与することができた. 付与できた映画数と記事名の例を感情毎に表 4 に示した.

13,088 件の映画に感情を付与できた一方で, 感情を付与できなかった映画も多数存在する. その理由として, 映画トピックに含まれていても映画を推薦する記事でないものが多く存在したことが挙げられる. 例えば「映画アナと雪の女王で使用された音楽 10 選」などは NAVER まとめの映画トピックに含まれているが, 映画ではない. また「おすすめ映画 20 選」などの印象表現が含まれていない記事に含まれる映画には感情を付与することができなかった.

また表 4 にあるように「怒」と「恥」の感情が付与できなかった理由として, その感情をもった印象語が「恥ずかしい」「照れくさい」「おもしろくない」「気持ちが悪い」など, 映画の印象を表すのに適していないものだったと考えられる. 「怒」の感情は 2 つの映画に付与できているが, 映画に関する記事ではなかったため, 記事名をのせていない. この結果から, 今後感情カラムの「怒」と「恥」は用いない. 本研究では, ユーザから推薦されている上位 100 件の映画を定番映画とし, そのうち上位 10 件を以下の表 6 に示す. 表 6 の 1 番上にある「ショーシャンクの空に」は最もユーザからの推薦数が多かった映画であり, その数が 63 であることを示している.

表 4　感情付与した記事の数とその例

感情	記事名 (例)	映画数
喜	元気が湧き出る映画まとめ	3,321
怒	-	2
哀	休日に見よう！絶対に泣ける映画 11 選	1,807
怖	最高に怖いおすすめのホラー映画まとめ【観覧注意】	2,301
恥	-	0
好	「あ～恋愛がしたい」胸キュン映画 10 選	3,516
厭	閲覧注意！エグい「グロテスク映画」30 作品！	1,107
昂	ラスト・どんでん返しがすごい！面白い映画 10 選	347
安	心の温まるストーリー映画 7 選！	347
驚	新事実に驚愕するドキュメンタリー映画 5 選	340

表 5　映画データベースの体系一例

title	movie	喜	哀	怖	好	厭	昂	安	驚
笑える恋愛映画まとめ	メリーに首ったけ	1	0	0	1	0	0	0	0
笑える恋愛映画まとめ	50 回目のファースト・キス	1	0	0	1	0	0	0	0
笑える恋愛映画まとめ	ラブ・アクチュアリー	1	0	0	1	0	0	0	0
号泣！せつない恋愛映画	バタフライ・エフェクト	0	1	0	1	0	0	0	0
号泣！せつない恋愛映画	(500) 日のサマー	0	1	0	1	0	0	0	0
号泣！せつない恋愛映画	ブルーバレンタイン	0	1	0	1	0	0	0	0
号泣！せつない恋愛映画	ブロークバック・マウンテン	0	1	0	1	0	0	0	0

表 6　ユーザからの推薦が多かった定番映画上位 10 件

順位	定番映画	件数
1	ショーシャンクの空に	63
2	ファイト・クラブ	55
3	時をかける少女	55
4	バタフライ・エフェクト	48
5	プラダを着た悪魔	47
6	ラブ・アクチュアリー	45
7	きみに読む物語	41
8	インセプション	40
9	タイタニック	40
10	セブン	38

4.　印象ベクトルに基づく映画推薦手法

本章では 3 章で明らかとなった定番映画の持つ印象から映画の推薦手法を提案し, 考察を行う. まず 4.1 章で印象ベクトルを用いた映画の類似度算出方法を述べ, 実際に定番映画と類似した映画を推薦できるかを検証する. 4.2 章では複数の定番映画を用いた映画推薦を行い, その結果に対して考察を行っていく.

4.1　印象ベクトルを用いた類似度算出方法

本論文では複数の印象を含む代表的な映画を例示する事で, それと類似した映画を推薦することを目指す. 第 3 章で述べたように, 表 5 のようなそれぞれの感情カラムに 1 もしくは 0 を付与した値が印象ベクトルであり, 映画の類似度を算出するためには, この印象ベクトルを用いる. これらの印象ベクトルの値を映画名毎に集計することで, それぞれの映画にどの印象が含まれているのかが分かる. 結果の定番映画上位 10 件を表 7 に示す.

表 7　定番映画と印象ベクトルを集計した数

映画名	喜	哀	怖	好	厭	昂	安	驚
ショーシャンクの空に	8	12	0	2	2	2	0	1
ファイト・クラブ	5	2	2	2	5	6	0	6
時をかける少女	1	6	0	6	0	0	0	0
バタフライ・エフェクト	6	5	1	8	4	3	0	3
プラダを着た悪魔	7	1	1	9	0	1	2	0
ラブ・アクチュアリー	3	8	0	23	0	0	0	0
きみに読む物語	1	12	0	20	0	0	0	0
インセプション	5	0	1	1	1	1	0	1
タイタニック	1	5	0	7	0	0	0	0
セブン	0	0	2	0	8	7	0	5

表 7 を見ると, 「ショーシャンクの空に」は, 「哀」と「喜」の感情が多く付与されており, 泣ける映画であるが, どこか明るい場面がある映画だと推測することができる. 「ラブ・アクチュアリー」では「好」の感情が最も多く付与されており, 次に「哀」の感情が多く付与されていることから, 泣けて愛のある映画であることが推測される. よって印象ベクトルの集計によりそれぞれの映画の印象が表れていることがわかる. 本節ではこの集計された印象ベクトルの値を用いて映画同士の類似度をはかる. 類似度の算出方法として, ここではコサイン類似度を用いる. 映画 i と映画 j の類似度を算出式を以下に示す. ここで $tp_{i,k}$ は映画 i の特徴ベクトルの k 番目の要素の値を表す.

$$cos_{tp}(i,j) = \frac{\sum_{k=1}^{K} tp_{i,k} * tp_{j,k}}{\sqrt{\sum_{k=1}^{K} tp_{i,k}^2 * \sum_{k=1}^{K} tp_{j,k}^2}}. \quad (1)$$

コサイン類似度を用いて定番映画との類似度を算出した結果, 定番映画と類似している映画, そうでない映画も示す事ができた. ここでは恋愛映画である「ラブ・アクチュアリー」を例に挙げ, この映画と類似度の高かった映画上位 10 件を表 9 に示し, 類似度の低かった映画上位 10 件を表 10 に示す. また Yahoo!映画のレビューに掲載されている, ユーザが映画を鑑賞して感じた代表的な印象を表すイメージワードとの比較を行い, 本研究の類似度が正しいかどうかを検証する. Yahoo!映画で扱うイメージワードは以下の通りである.

表 8　Yahoo!映画のイメージワード一覧

泣ける, 笑える, 楽しい, 悲しい, ファンタジー, スペクタクル, ゴージャス, ロマンチック, 不思議, パニック, 不気味, 恐怖, 勇敢, 知的, 絶望的, 切ない, セクシー, かわいい, かっこいい, コミカル

表 9　ラブ・アクチュアリーと類似度の高い映画上位 10 件

映画名	類似度	イメージワード
ゴースト ニューヨークの幻	0.990	泣ける, ロマンチック, 切ない
ノッティングヒルの恋人	0.988	ロマンチック, 楽しい, 笑える
エターナル・サンシャイン	0.981	切ない, ロマンチック, 不思議
(500) 日のサマー	0.980	切ない, かわいい, ロマンチック
きみに読む物語	0.975	ロマンチック, 泣ける, 切ない
ホリデイ	0.971	ロマンチック, 楽しい, かわいい
シザーハンズ	0.963	切ない, ファンタジー, 泣ける
ジョゼと虎と魚たち	0.960	切ない, 泣ける, 悲しい
タイタニック	0.960	泣ける, ロマンチック, 切ない
50 回目のファースト・キス	0.953	ロマンチック, 楽しい, 笑える

表 10　ラブ・アクチュアリーと類似度の低い映画上位 10 件

映画名	類似度	イメージワード
アフタースクール	0.017	楽しい, 笑える, コミカル
シャイニング	0.023	恐怖, 不気味, パニック
ジョーズ	0.028	恐怖, パニック, 勇敢
ユージュアル・サスペクツ	0.03	知的, かっこいい, 不気味
メメント	0.031	知的, 不思議, 絶望的
シックス・センス	0.047	切ない, 泣ける, 絶望的
羊達の沈黙	0.054	恐怖, 不気味, 知的
ミスト	0.059	絶望的, 恐怖, パニック
シャッター・アイランド	0.068	不気味, 絶望的, 切ない
バイオハザード	0.073	かっこいい, 勇敢, 恐怖

表 9 を見ると「ラブ・アクチュアリー」と類似度の高い上位 10 件の映画には Yahoo!映画のイメージワード,「ロマンチック」や「泣ける」などが多く含まれており, 本研究で「好」の感情や「哀」の感情が多く付与された「ラブ・アクチュアリー」と近いジャンルの映画が上位に存在していることがわかる. 表 10 を見ると表 9 に頻出していた「ロマンチック」のイメージワードとは異なり,「恐怖」「不気味」「絶望的」といったイメージワードが目立つ. このことから本研究の「好」や「泣」の感情とは離れており,「怖」や「厭」の感情と近いイメージワードが並んでいるため類似度が低い映画であると言える. この結果により定番映画からそれと類似した映画を推薦することは可能である.

4.2　複数の定番映画を用いた映画の推薦

前節ではコサイン類似度を用いて, 定番映画と類似度の高い映画を推薦できることを示した. 本節でも同じように定番映画との類似度をコサイン類似度で算出して, 類似度の高い映画を推薦するが, 今回は 1 つの定番映画から新たな映画を推薦するのではなく, 複数の定番映画を組み合わせることで, 特徴的な印象を持った映画の推薦を行う. 方法として, 複数の定番映画における集計された印象ベクトルの最大値をとることで, 映画の特徴的な印象を取り出し, 複数の映画と似た新たな映画を推薦する. 例えば「タイタニック」と「テッド」を組み合わせた場合,「タイタニック」と「テッド」の特徴を持つ映画を推薦することができる. 上記の組み合わせの印象ベクトルと似た映画を, 類似度の高かった順で上位 11 件の映画を表 11 に示す.

表 11　「タイタニック」「テッド」の組み合わせと類似度の高い映画上位 10 件

映画名	類似度	イメージワード
フラガール	0.953	泣ける, 楽しい, 笑える
スタンド・バイ・ミー	0.950	切ない, 泣ける, 楽しい
メリーに首ったけ	0.937	笑える, 楽しい, かわいい
ギルバート・グレイプ	0.936	切ない, 泣ける, 悲しい
エリン・ブロコビッチ	0.936	かっこいい, 勇敢, セクシー
プラダを着た悪魔	0.901	ゴージャズ, かわいい, 楽しい
リトル・ミス・サンシャイン	0.888	笑える, 楽しい, コミカル
下妻物語	0.884	笑える, 楽しい, かわいい
ヘアスプレー	0.868	楽しい, かわいい, 笑える
50 回目のファースト・キス	0.868	ロマンチック, 楽しい, 笑える

「タイタニック」に最も多く付与されている感情は「好」で,「テッド」に最も多く付与されている感情は「喜」である. この 2 つの映画の組み合わせにより, それぞれの最大値をとるので「好」と「喜」の感情が強く現れる映画が推薦される. 表 11 を見ると「好」の感情と似たものとして「かわいい」のイメージワードが含まれており, また「喜」の感情と似た「楽しい」「笑える」「コミカル」などのイメージワードも多く含まれている. この結果より定番映画の組み合わせによって特徴的な感情を合わせ持つ映画を推薦できると言える.

5. おわりに

5.1　まとめ

本論文では,「監督名」や「キャスト名」などの属性情報を用いて未知の映画タイトルを探す事が困難であること, ユーザによって印象に個人差があるという 2 つの問題を解決し, 映画に関する印象が複数含まれている定番映画を例示することで新たな映画推薦を目指す. そのため映画を鑑賞して得た印象が掲載されている記事タイトルと, 映画が整理・体系化されている NAVER まとめを用いた. その中から多くの印象が含まれている定番映画を定義することできた.

まず定番映画を見つけるために, NAVER まとめから 6,564 件の記事タイトルと 52,500 件の映画を取得を行い, 記事名, 映画名, 感情の 3 要素を設けた映画データベースに全データを格納した. 記事タイトルから印象を抽出するために, 映画の印象を表す印象語を用いて独自の印象辞書を生成した. その後, 取得した映画名の修正を行い, ユーザが推薦した数の多い上位 100 件の映画を定番映画とした. また定番映画にどんな感情が含まれているかを示すために, 中村の感情辞典の中から「喜・哀・怖・好・厭・昂・安・驚」の 8 つの感情軸を用いて, 定番映画に印象ベクトルを付与した. この印象ベクトルを集計することで, 定番映画に感情がどの程度付与されているかが分かり, 映画の特徴を数値で表すことが出来た. 次にこの印象ベクトルが付与された定番映画と類似度の高い映画を示すために, コサイン類似度を用いて類似度算出を行った. そこで本研究の類似度が正しいかを検証するために, Yahoo!映画のイメージワードとの比較を行った. その結果, 本研究で付与した感情を持つ定番映画と高い類似度の値を示した映画には, 本研究の感情と近いイメージワードが付与されていた. また 1 つの定番映画から新たな映画

を推薦するだけでなく, 複数の映画を組み合わせることで, より特徴的な印象を持った映画を推薦することができた.

5.2 今後の課題

今後の課題としては, 複数の定番映画の組み合わせによる新たな映画の推薦は, 類似度の値が高い映画でも, 本当にそれぞれの定番映画の特徴を捉えているとは言えないのではないかという結果になることもあった. 原因として定番映画の中には付与された感情数が少ないものが多く含まれていることが挙げられる. この問題を解決するためには, 感情を付与できるよう映画により合った印象語の充実や, 感情の付与できなかった映画にも感情を付与することで, 感情の付与数を増やし, 類似度の値を確かにしていくべきと考える.

6. 謝　　辞

本研究の一部は, JSPS 科研費 25280110 の助成を受けたものです.

文　　献

[1] 小野 智弘, 黒川 茂莉, 本村 陽一, 麻生 英樹. ユーザ嗜好の個人差と状況差を考慮した映画推薦システムの実現と評価. **情報処理学会論文誌**, Vol. 49, No. 1, pp.130-140, 2008.

[2] Miller B.N., Albert I., Lam S.K., Konstan J.A., and Riedel J. Movielens unplugged:experiences with an occasionally connected recommender system. *Proc. 8th Int. Conf on Intelligent User Interfaces*, pp.263–266, 2003.

[3] 上松 陽介, 疋田 輝雄. Linked data を用いたユーザの興味に基づく映画推薦手法の提案. **情報処理学会**, **全国大会講演論文集**, pp.585–587, 2013.

[4] R. Plutchick. The nature of emotions. *American Scientist*, Vol.89, pp.344-355, 2011.

[5] 中村明. **「感情表現辞典」**. 東京堂出版, 1993.

[6] 熊本 忠彦, 河合 由起子, 田中 克己. 新聞記事を対象とするテキスト印象マイニング手法の設計と評価. **電子情報通信学会論文誌**, Vol. J94-D, No. 3, pp.540-548, 2011.