

# 映画レビューの印象タグの希少性に基づく協調フィルタリング

北中 雄也<sup>†</sup> 北山 大輔<sup>†</sup>

<sup>†</sup> 工学院大学情報学部コンピュータ科学科 〒163-8677 東京都新宿区西新宿 1-24-2

E-mail: [†j112035@ns.kogakuin.ac.jp](mailto:†j112035@ns.kogakuin.ac.jp), [kitayama@cc.kogakuin.ac.jp](mailto:kitayama@cc.kogakuin.ac.jp)

あらまし 一般的な協調フィルタリングによる推薦システムでは、ユーザ評価を用いて類似ユーザを決定し、類似ユーザに基づき推薦アイテムを決定する。しかしながら、映画のように人によって感じ方が異なるコンテンツに対しては、類似ユーザを決定する際にユーザ評価のみならず印象を加味して推薦アイテムを決定する必要がある。実際、映画サイトである映画.comには、映画に対し印象タグがつけられるようになっており、映画に対し、印象は重要であるといえる。そこで我々は、ある映画を見た際の印象が似た他のユーザを探し、そのユーザがレビューをつけた映画を推薦することで、より満足度の大きい映画を探せるのではないかと考えた。具体的には、ある映画に対して、一般につけられる印象タグではない希少な印象タグほど、ユーザの感じ方を強く表現していると考え、このようなタグに高い印象度を与える。印象度が高いタグが一致している他のユーザは、より映画を見る際の感じ方が近いと考え、そのユーザに関しては類似度を上げる。ユーザ間で同じ映画に関して印象の類似度を算出し、その合計値の高いユーザが高い評価をつけた映画を推薦する手法を提案する。

キーワード 映画推薦, 印象タグ, 協調フィルタリング

## 1. はじめに

近年映画.com<sup>(注1)</sup>やyahoo映画<sup>(注2)</sup>などの映画のレビューサイトがさかんであり、ユーザは視聴する映画を決める際にレビューサイトを参照することがある。そのようなレビューサイトでは、ユーザはレビューと共に印象のキーワードを投稿することが可能である。また、レビューサイト側でも投稿された印象の数をもとに、映画のメタデータとして印象を決定している。図1の映画の情報はyahoo!映画のアメリカン・スナイパーである。ユーザレビュー中での投稿頻度の上位である「悲しい」、「切ない」、「勇敢」が代表的な印象となっている。現在の映画推薦は、図2のようにユーザの映画に対する評価値を用いて協調フィルタリングを行い、その映画に対する評価値の近いユーザを類似ユーザと決定し映画を推薦するものや、ユーザの評価値や印象をコンテンツの特徴として用いよく似た映画を推薦するものもある。協調フィルタリングによる推薦では類似ユーザの決定にユーザたちの印象を考慮しておらず評価値のみを使い映画を推薦するのが一般的である。また後者のコンテンツベースの推薦は映画に対する印象を特徴として利用しているが、映画のメタデータとしての印象は多数のユーザがつけたものを用いるのが一般的である。そのため、これらの推薦は個人によって異なる映画を観たときの印象が考慮されていないため、推薦された映画を観たユーザが思っていた印象と異なるときに満足度が低くなる場合がある。そこで我々は、個人の印象タグの付け方から、それに似た印象タグの付け方をしているユーザを探しそのユーザが観た映画から映画を推薦する手法を提案する。本論文の構成を以下に示す。まず2節で本研究のアプローチにつ



図1 映画の印象の例

いて、3節では印象度に基づく映画推薦について、4節ではプロトタイプシステムについての構成と予備実験について、5節では本手法の評価実験について述べる。6節ではまとめと今後の予定についてを述べる。

## 2. 本研究のアプローチ

### 2.1 映画レビューサイトと印象タグ

本研究では、映画.comにある映画レビューを用いてユーザの付けた印象タグおよび評価をメタデータとして扱う。映画.comは1998年より開設されて以来日々映画情報の数を増やしている。現在レビュー数で最も多いものは619件ものレビューが付けられており、各レビューには表1中の12種類の印象タグから付けられる。また、映画.comの印象タグは一つの映画に対して最も少なくても0個、最も多くても3個の印象タグを付けることが可能である。その他には、評価値を最大で5.0までの数値で表しレビュー文を掲載することが可能である。

(注1) : <http://eiga.com/>

(注2) : <http://movies.yahoo.co.jp/>

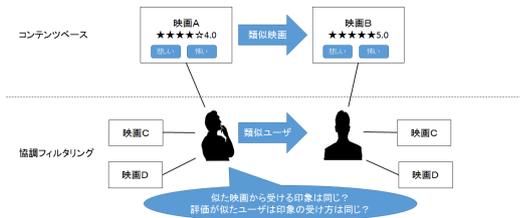


図 2 従来の推薦の問題

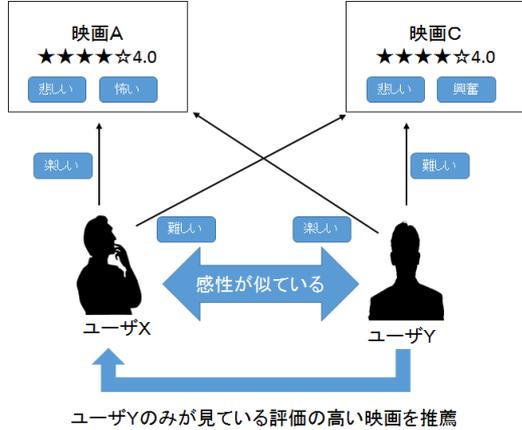


図 3 提案手法の概要

表 1 印象タグ

難しい	萌える
興奮	笑える
知的	泣ける
楽しい	悲しい
怖い	幸せ
寝られる	単純

## 2.2 提案手法の概要

映画のレビューで、一般につけられる印象タグではない希少な印象タグほど、その印象タグをつけたユーザの感じ方が強く現れていると考え、このような印象タグに高い印象度をつける。我々は印象度が高いタグが一致している他のユーザは、より自身と映画を見る際の感じ方が近いのではないかと考え、そのユーザに関して類似度が高くなるように算出する。自身と他ユーザの間で同じ映画に関して印象タグの類似度を算出し、その合計値の高いユーザが高い評価をつけた映画を推薦する手法である。例えば、図 3 では映画 A の一般の印象は楽しいであるがユーザの中には悲しいと感じたユーザ X もおり、そのようなユーザに対して同じように映画 A に対して悲しいと感じたユーザ Y がいる。手法により、ユーザ Y の感性をユーザ X と似ていると判断し、ユーザ Y の観た映画 C を推薦しようとしている。

## 2.3 関連研究

映画につけたレビューの評価から個人の嗜好の判断を行い高評価の映画をユーザの求めているものとして近い映画を推薦する手法 [1] [2] や、ユーザがどのような印象を好んでいるかを推定しその印象が付けられた映画を推薦する研究 [3] [4] [5]、またユーザの評価から他ユーザとの類似度を測り映画を推薦する手法もある [6]。他には個人の嗜好を判断し推薦するコンテンツの

表 2 映画 A の印象度

印象	タグ数	印象度
萌える	1	8.97
笑える	2	7.97
寝られる	2	7.97
楽しい	3	7.38
幸せ	4	6.97
単純	14	5.16
泣ける	35	3.84
知的	39	3.68
難しい	56	3.16
興奮	72	2.80
怖い	104	2.27
悲しい	169	1.57

決定に役立てる研究 [7] [8] [9] や、多くの人が評価したものの数値を下げあまり評価が付いていないものの数値を上げることににより、多く評価が付いたものがより推薦対象になりやすいという問題を解決している研究もある [10]。そして推薦する物を選ぶ際に、一般的な嗜好ではなく他ユーザの嗜好を反映させて、同じようなものばかりが推薦されないようにするための協調フィルタリングの研究もある [11]。これらの研究は、いずれも推薦対象を選ぶ際に多くのユーザが付けた印象がそのまま推薦対象の印象として決定しているが、本手法では、個人の印象がそのまま推薦対象の印象になる推薦手法である。

## 3. 印象度に基づく映画推薦

### 3.1 印象タグの印象度

各映画に付けられている印象タグのうち多くのユーザがつけている印象タグは、個人の感性を特徴付けているとはいいがたい。そこであまり付けられていない希少な印象タグは個人の感性の特徴が強くと考え、上記の考えに基づき映画ごとの印象タグに印象度 *impression* を算出する。印象度の算出にはシャノンの情報量を用い、希少性のあるタグほど印象度が高い値になる。

$$impression(t_k, T) = -\log_2 \frac{|t_k|}{|T|} \quad (1)$$

ここで、 $t_k$  はある種類  $k$  の印象タグの集合であり、 $T$  はある映画につけられたすべての印象タグの集合である。なお、 $t_k \subset T$  である。すなわち、 $\frac{|t_k|}{|T|}$  は、ある映画における  $t_k$  の付与確率である。印象度の算出例は表 2 である。

### 3.2 ユーザの類似度算出と映画推薦

類似度 *sim* を算出するには一致タグの印象度を算出した後に一致しなかったタグの印象度を減算する。これを各映画ごとに行い類似度を加算していくことにより自身と他ユーザの類似性スコアを表す。ただし類似度が 0 以下の場合には類似性スコアに 0 を加算する。ユーザ X の映画 A に対するユーザ Y, Z の類似度算出例を図 4 とする。

$$sim(X_a, Y_a) = com(X_a, Y_a) - ncom(X_a, Y_a) \quad (2)$$

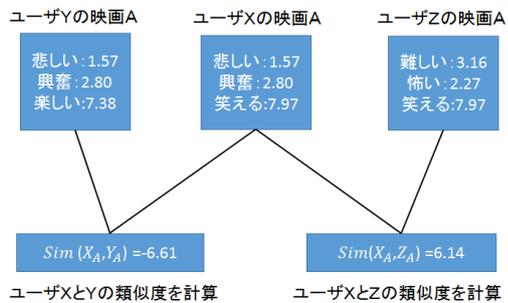


図 4 類似度算出例

$$score(X, Y) = \sum_{i=1}^n \delta(sim(\mathbf{X}_i, \mathbf{Y}_i)) \quad (3)$$

ユーザ  $X, Y$  が観た映画  $a$  の類似度を  $sim(X_a, Y_a)$  とし、ユーザ  $X, Y$  が観た映画  $a$  の一致したタグの印象度の合計を  $com(X_a, Y_a)$  と表し、ユーザ  $X, Y$  が観た映画  $a$  の一致していないタグの印象度の合計を  $ncom(X_a, Y_a)$  で表す。  $n$  は一致した映画の数を表す。  $\delta$  は数値が 0 以下の場合に 0 を返す関数である。このようにして算出した  $score$  の高いユーザが高い評価値をつけた映画を推薦する。

#### 4. プロトタイプシステム

映画.com から収集した映画情報やユーザ情報をシステムのデータセットとする。収集したユーザ 18,581 件のユーザ ID から入力するとそのユーザ ID が評価をつけた映画集合を抽出し、抽出した映画に評価をつけたユーザを抽出する。抽出したユーザの映画集合から入力したユーザと共通する映画の印象タグを用いユーザ間の類似性スコアを算出する。そこで算出された類似性スコアからユーザを順位付けを行い、上位のユーザが高評価をつけた映画を推薦する。図 5 は、ユーザ間の類似度算出部のフローである。推薦結果は図 6 のように表示される。ユーザ 1, ユーザ 2 は類似性スコア上位のユーザであり、そのユーザの評価値が高い映画を順に表示するものとなっている。

#### 5. 評価実験

##### 5.1 類似度算出の評価

###### 5.1.1 実験方法

一人のユーザ情報を対象としてそのユーザ情報と一つ以上同じ映画を観ているユーザ情報を無作為に選び、一致映画数、一致タグ数、類似性スコアを調べ順位付けをする。各順位付けでどのような差が出たかを考察し本手法である印象タグの希少性を考慮した類似度が有用であったかを確認する。

###### 5.1.2 結果と考察

表 3 は、左からユーザ、一致映画数、一致タグ数、類似性スコアとなっており括弧内の数字は各指標による順位となっている。レビュー数 127 件、総タグ数 246 件のユーザの情報を対象とした結果、ユーザ 1 に関して一致映画数、一致タグ数は 1 位ではないが、類似性スコアでは 1 位となった。従来の指標に基づき、一致映画数、一致タグ数を見た結果では、ユーザ 2 がどの項目においても高い数値を出しトップとなっているが、類

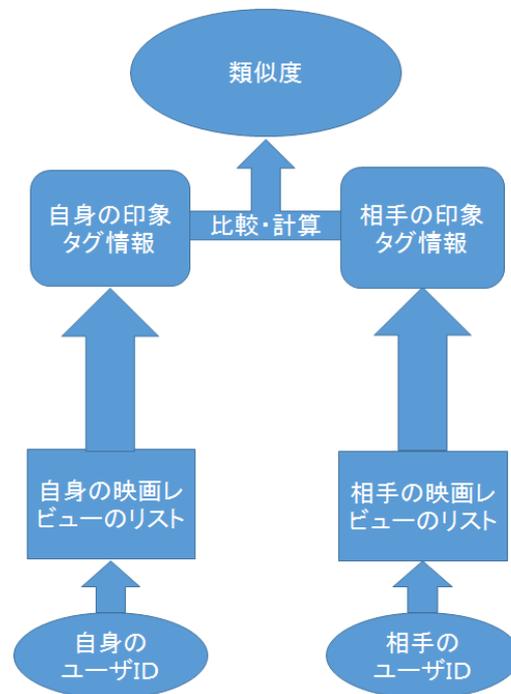


図 5 システムの流れ

ユーザ1	 ・評価・レビュー 4.0 ★★★★★ (全188件)	 ・評価・レビュー 4.0 ★★★★★ (全585件)	 ・評価・レビュー 3.7 ★★★★★ (全111件)
ユーザ2	 ・評価・レビュー 4.2 ★★★★★ (全314件)	 ・評価・レビュー 4.0 ★★★★★ (全144件)	 ・評価・レビュー 3.8 ★★★★★ (全324件)

図 6 推薦結果の例

似性スコアの結果を見ると希少タグが一致したほうが類似性スコアが高くなっているのがわかる。またこれら以外の類似性スコアの順位を見ても一致タグ数や一致映画数とは、順位が異なり、独立した指標であることを確認した。これらの結果から一致タグ数が多いだけでなく希少性の高いタグが一致しているほど本システムでは類似度が上がりよりユーザ自身との個性に近いユーザがわかる。

##### 5.2 推薦精度の評価

###### 5.2.1 実験方法

18581 個のユーザの視聴映画、映画レビューの集合から無作為に選択したユーザの映画情報集合 12 件を対象とした。選択

表 3 類似性スコアの高いユーザ上位 10 件

ユーザ	一致映画数	一致タグ数	類似性スコア
ユーザ 1	35(4)	41(2)	68.01(1)
ユーザ 2	77(1)	56(1)	53.84(2)
ユーザ 3	22(8)	17(9)	39.80(3)
ユーザ 4	21(9)	20(6)	38.85(4)
ユーザ 5	27(6)	26(5)	38.09(5)
ユーザ 6	61(2)	41(2)	37.64(6)
ユーザ 7	21(9)	19(7)	29.28(7)
ユーザ 8	6(21)	6(16)	28.04(8)
ユーザ 9	50(3)	35(4)	20.55(9)
ユーザ 10	35(4)	18(8)	17.33(10)

した集合から無作為に一つの映画の情報を削除し、映画情報が一つ抜けた状態で類似するユーザの映画情報集合上位 10 件を抽出した。抽出した集合に含まれる削除した映画を用いて評価する実験を行った。11 件は評価値が高いものを削除対象とし、1 件は評価値が低いものを削除対象とした。

### 5.2.2 結果と考察

評価値の高い映画を削除した場合の結果を示したものが表 4 である。一方、低い映画を削除した場合の結果を示したものが表 5 である。表は左から実験番号、対象の削除映画の評価値、全ユーザの削除映画に対する評価値の平均値、算出された集合中の削除映画の評価値の平均値、算出された集合中の削除映画の印象と対象の集合中の削除映画の印象の一致度を示す。11 件中 8 件が評価値を見ると全体平均から対象の評価値に 0.1 以上近づく結果となった。一方、実験 12 では低い評価値のほうにも近づくという結果が出た。印象の一致率では、評価値の高いものは、53%一致する結果となった。この結果から、似た印象をつけるユーザは他の映画でも似た印象をつけやすいことがわかる。評価値が低い映画の実験では一致するタグはなかった。この結果は評価値が低いものには上位 10 件のうち 4 件にしか印象を付けられていなかったという問題が原因であると考えた。実験の結果より本手法による推薦は評価値、印象が対象に近いものを推薦できることを確認した。

## 6. おわりに

本論文では、映画に付けられる印象タグのうち一般的ではない印象タグに個人の感じ方が強く現れていると考え、印象度と類似度を定義した。一致映画数や一致タグ数の順位付けと比べて本手法で算出した類似性スコアによる順位付けの有用性を示し、また本手法で推薦する映画は対象に近い評価値、印象を持つものが推薦されるという有用性を示した。今後の課題として実験数を増やし、より詳細なデータを取ることやレビューの少ないユーザが推薦対象となりにくい問題の対処が必要となる。また評価値の低い映画の印象一致の精度向上も視野に入れていく必要があり、一致する映画数と精度の関係を明らかにしていく必要もある。また、本稿では、映画に対する希少性しか用いていないが、ユーザが普段付けないタグがユーザの印象を表すというアプローチもあるので、それを加味した印象度の算出も今後の課題である。

表 4 評価値の高い映画での評価実験

実験番号	対象の評価値	全体の平均評価値	推薦の平均評価値	印象一致率
実験 1	5.0	3.5	3.4	73%
実験 2	5.0	4.0	4.0	45%
実験 3	5.0	3.9	4.4	61%
実験 4	5.0	3.6	3.7	46%
実験 5	4.5	2.8	3.2	23%
実験 6	4.5	3.9	4.1	78%
実験 7	5.0	3.7	3.9	60%
実験 8	5.0	2.8	3.4	17%
実験 9	5.0	3.7	3.7	73%
実験 10	5.0	3.9	4.1	44%
実験 11	5.0	3.6	3.6	63%
平均	4.9	3.6	3.8	53%

表 5 評価値の低い映画での評価実験

実験番号	対象の評価値	全体の平均評価値	推薦の平均評価値	印象一致率
実験 12	0.5	4.0	3.7	0%

## 謝 辞

本研究の一部は、平成 27 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] 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.
- [2] 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. 情報処理学会第 76 回全国大会講演論文集, Vol. 2014, No. 1, pp. 619-620, 2014.
- [3] 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.
- [4] 上松輝介, 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. 情報処理学会第 75 回全国大会講演論文集, Vol. 2013, No. 1, pp. 585-586, 2013.
- [5] 智弘小野, 陽一本村, 英樹麻生. ベイジアンネットによる映画コンテンツ推薦方式の検討. 電子情報通信学会技術研究報告. NC, ニューロコンピューティング, Vol. 104, No. 348, pp. 55-60, 2004.
- [6] 慎平岡本, 基衛東. 適合性フィードバックを利用した映画推薦システムの研究. 情報処理学会第 71 回全国大会講演論文集, Vol. 71, pp. 577-578, 2009.
- [7] 智弘小野, 茂莉黒川, 陽一本村, 英樹麻生. 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.
- [8] 大弘吉川, 貴章森, 武古橋. Personalizability を考慮した推薦システムの提案. 情報処理学会論文誌数理モデル化と応用 (TOM), Vol. 6, No. 1, pp. 111-118, 2013.
- [9] 晃一井口, 嘉徳土方, 正吾西田. ユーザプロファイルの仮想的個別化によるテレビ番組推薦. 人工知能学会論文誌, Vol. 30, No. 1, pp. 71-83, 2015.
- [10] 和裕北川, 範子新井. 情報量規準を用いた協調フィルタリング. 日本計算機統計学会大会論文集, No. 19, pp. 17-20, 2005.
- [11] 伊東孝浩, 加藤昇平. ユーザ評価と楽曲ゆらぎ特徴を用いた楽曲嗜好のハイブリッド推定—共有楽曲数の差異が協調フィルタリングに与える影響—. 研究報告知能システム (ICS), Vol. 2014, No. 14, pp. 1-5, 2014.