

# 内容に即した推薦文を提示する映画推薦システム

石田 雄登<sup>1</sup> 打矢 隆弘<sup>1</sup> 内匠 逸<sup>1</sup>

**概要:** 近年、インターネットの普及により様々な商品を取り扱う EC サイトが増加してきた。しかし、膨大な数の商品の中から自分の好みに合う商品を探し出す事はユーザにとって作業負担が大きい。この問題を解決するために、ユーザに対して「おすすめのアイテム」を提示する推薦システムの開発及び導入が進められている。現在、Web サイト上で導入されているほとんどの推薦システムは具体的な推薦理由の提示を行わない。推薦されたアイテムの中からユーザがアイテムを選択する際、電化製品等であれば性能による選択が可能である。一方、本や映画、嗜好品等のアイテムの場合、ユーザの嗜好の影響が大きいため性能による選択が困難である。

この問題を解決するために、ユーザの意思決定要因として推薦理由の提示を行う。本研究では、Web 上のレビューを利用して、推薦対象の内容に即した推薦理由の作成を行う。また、ユーザの嗜好を抽出し、推薦理由の作成に利用する。これにより、パーソナライズされた推薦理由の作成を可能とする。

また、本研究ではアンケートを用いて作成した推薦文を提示する映画推薦システムの評価を行った。実験の結果、推薦文を提示しない状態で被験者が興味を持たなかった映画に対して、推薦文を提示することで被験者の興味を喚起する傾向を確認した。これにより、内容に即した推薦文のアイテム選択要因としての有効性を確認した。

## Movie Recommendation System to Showing the Reason Suitbale for Content

YUTO ISHIDA<sup>1</sup> TAKAHIRO UCHIYA<sup>1</sup> ICHI TAKUMI<sup>1</sup>

### 1. はじめに

現在、多くの Web サイトにおいて情報推薦システムが導入されている。情報推薦システムとは、インターネットを利用するユーザに対して適切で有益な情報を提示する情報収集支援システムである。情報推薦システムは、ユーザが行う情報選択を自動化し、個人のニーズに合った情報を提供することを目的としている。ユーザ個人のニーズを把握するために、一般的に情報推薦システムはユーザの行動からユーザの嗜好を取得し、ユーザプロフィールを作成する。そして、そのユーザプロフィールを何らかの基準と比較し、ユーザが個々のアイテムにつけるであろう評価を予測し、評価値が高いアイテムを推薦する。

この情報推薦システムの実用例として、Amazon 等の EC<sup>\*1</sup> サイトが挙げられる。このような EC サイトでは、

ユーザの検索履歴や購買履歴からユーザプロフィールを作成し、ユーザが高く評価値をつけると予測されるアイテムを推薦する。これにより、ユーザは自らの嗜好に近い商品や必要とする商品を検索する手間を削減する事が出来る。また Renjie ら [1] は、動画配信サービス YouTube<sup>\*2</sup> における協調フィルタリングに基づいた情報推薦システムの有用性を検証している。この検証では、人気ランキングを表示する場合と比較して、提示された動画のクリック率が 2.07 倍向上したと報告されている。これは、情報推薦システムがユーザの Web における回遊性を向上させることが可能であり、より多くのアイテムをユーザに閲覧させる効果があるといえる。

ユーザは推薦システムが提示するアイテム群から必要なアイテムや好みのアイテムを選択する。この際、アイテムの性能やユーザ評価から総合的に判断してアイテムの選択を行う。しかし、映画や音楽などユーザの嗜好の影響が強

<sup>1</sup> 名古屋工業大学 大学院工学研究科

<sup>\*1</sup> electronic commerce

<sup>\*2</sup> <https://www.youtube.com>

く反映されるアイテムの場合、アイテムの性能で比較を行うことが困難である。ユーザは、このようなアイテムの選択を行う際、他のユーザ評価を閲覧し、自らの嗜好に合っているかを判断する必要がある。ユーザに提示するアイテムが多くなるほどこの作業の負担は大きくなる。

この問題を解決するために、私達は、推薦アイテムに対して推薦理由を付与するシステムの提案を行う。本研究では、推薦アイテムの内容で特にユーザにとってアイテム選択要因となり得ると考えられる文章を推薦理由として提示する。これにより、性能でアイテムの選択が出来ない場合において有用なアイテム選択要因の提示を目指す。

しかし、アイテムの作成者が推薦アイテムの内容を事前知識としてコンピュータに学習させるためのデータを作成することはアイテム作成者の負担となる。本研究では、推薦アイテムの内容を推薦アイテムに対するレビューから取得する。これにより、アイテム作成者に負担をかけないようにする。

本論の構成は以下の通りである。第2章では情報推薦システムについての概要と、情報推薦の手法について述べる。第3章では関連研究に対する考察について述べる。第4章では本システムの設計について述べる。第5章では提案システムの試作と実験について述べる。第6章ではまとめと今後の予定について述べる。

## 2. 情報推薦システム

情報推薦システムは、インターネットを利用するユーザに対して適切で有益な情報を提示する情報収集支援システムである。ユーザの情報収集支援アプローチの一つである情報検索は、ユーザのクエリ入力に基づいて情報獲得を行う。これに対して情報推薦システムはユーザの能動的な行為が無くても情報を獲得する事が可能であるという特徴がある。

### 2.1 推薦アイテムの選択手法

ユーザに対して推薦するアイテムの選択手法は、各々が特徴を持っている。ここでは代表的な二つの選択手法である、「協調フィルタリング」と「内容ベースフィルタリング」の説明を述べる。

現在最も主流である協調フィルタリングは、アイテムを推薦する対象ユーザと似た嗜好を持つユーザの好むアイテムを推薦する手法である。協調フィルタリングの利点として、メタデータの準備が不要であるという点やアイテムのジャンルに関係なく推薦を行う事が出来るため汎用性が高いという点があげられる。これは、協調フィルタリングはアイテムの内容や性質に依存しないためである。

一方で、内容ベースフィルタリングは、アイテムを推薦する対象ユーザの嗜好と似た性質を持つアイテムを推薦する手法である。内容ベースフィルタリングの利点として、

利用者が少ない状況でも運用可能であるという点やユーザの嗜好を直接利用したアイテムの選択が可能であるという点があげられる。これは、内容ベースフィルタリングはアイテムの内容や性質を利用するためである。

提案システムでは、推薦理由として内容に即した推薦文を提示するため、アイテムの内容を利用した推薦アイテムの選択を行うほうが望ましい。よって、本研究では内容ベースフィルタリングを用いて推薦アイテムの選択を行う。

### 2.2 内容ベースフィルタリング

内容ベースフィルタリングは、ユーザ嗜好の特徴を記述したユーザプロフィールと、各アイテムの特徴から類似度を分析・計算し、類似度が高いアイテムを推薦する。例えば、本をアイテムとして扱う場合、本には著者やジャンルなどの付加的な情報が存在する。これらの情報を特徴と呼ぶ。また、特徴がアイテムをどの程度表現しているかを定量的に表した値を特徴量と呼ぶ。アイテムの特徴量は、テキストの場合にはキーワードの出現頻度などで表される。音楽データや映像データなどのメディアコンテンツの場合は、周波数成分・色彩の情報などで表される。抽出した特徴量はモデル化され、コンテンツモデルと呼ばれる。また、ユーザからは、アイテムに対する評価やアンケート、コンテンツの利用履歴等からコンテンツの特徴量に関する嗜好情報を抽出し、モデル化する。これがユーザプロフィールとなる。

### 2.3 情報推薦システムの評価指標

情報推薦システムにおいて評価指標はシステムの質と性能を判断するための重要な要素である。これまでに提案されてきた指標の多くは、推薦されたアイテムがユーザの嗜好に合っているかどうかを表す「推薦精度」に関するものである。しかし、近年では他の観点、例えばユーザが未知のアイテムを推薦する「新規性」や、ユーザにとって掘り出し物と感じるアイテムを見つける「意外性」などに関する評価指標も提案されている [2]。

## 3. 関連研究

本研究は、情報推薦の際にアイテムに関する付加情報を提示することによってシステムの利便性の向上を目的としている。本章では、本研究と関連した内容の研究について述べる。

### 3.1 推薦理由を提示する推薦システム

折原ら [3] は、ユーザ満足度と自己決定感の相関 [4] に注目し、ユーザがアイテムを選択する際に「何故このアイテムを選択したのか」を自らで理解し決定させる (自己決定理論) 事でユーザ満足度の向上を目指している。この研究では、Web ニュースを推薦アイテムとし、ユーザ満足度の

評価指標として NPS<sup>\*3</sup> を採用している。

提案手法としては、ユーザに対して情報を推薦する際、何故その情報が推薦されたのかを明示的にする為に、推薦する情報に含まれるキーワードのうち、ユーザが興味をもつキーワードを推薦理由として提示する。折原らはこのような推薦理由の明確化によってユーザ満足度の向上を図る事が可能かを検討する為の実験を行っている。

ユーザアンケートを用いた実験の結果、推薦理由を付与する事で、付与しない場合と比べてユーザ満足度が上昇する傾向があるとしている。これにより、キーワードを用いた推薦理由はユーザにとってアイテム選択要因として機能していると考えられる。一方で、提示された推薦理由の精度が低い事が問題点としてあげられている。よりの確な推薦理由を提示する事が出来ればさらなる満足度の向上が期待できる。

### 3.2 Web 上のレビューを利用した映画推薦システム

林ら [5] は、ユーザが投稿したレビューを基に類似する特徴を持つレビューを検索し、そのレビューが付与されている映画を推薦するシステムを提案している。

提案手法では、ユーザが投稿したレビュー内で肯定的な表現あるいは否定的な表現で評価されている単語をユーザの興味語として抽出している。この興味語からユーザプロフィールを作成する。同様の手法で各映画に対して投稿されているレビューの特徴を抽出し、ユーザプロフィールと類似度が高いレビューが投稿されている映画をレビュー文と共に提示する。林らはこの推薦手法の推薦精度やレビュー文による映画に対するユーザの印象の変化を知るための実験を行っている。

ユーザアンケートを用いた実験の結果より、推薦精度はランダム手法よりは高い値を示したが、*tf·idf* によって類似するレビューが投稿されている映画を推薦する従来手法との有意差は無かったとしている。一方で、ユーザに対して提示したレビュー文中の興味語は、映画に対するユーザの興味を喚起させる傾向があるとしている。

## 4. 提案システムの設計

本研究では、推薦の際にユーザの嗜好に合ったアイテムに関する推薦文の提示を行うことで、ユーザにアイテム選択の動機を与える事を目指す。

### 4.1 提案システムの概要

3章において述べたように、情報推薦の際に推薦理由を提示する事は、ユーザにとってアイテム選択の動機となると考えられる。一方で、キーワードを用いた推薦理由だけでは内容の把握が困難である。例えば、「子供」というキ

ワードで映画を推薦された場合、「子供向け」の映画か「子供が主人公の映画」かの判別は付けられない為、推薦理由としては適切ではない。そこで、本研究ではより内容に即した推薦理由の提示を行う。まず、実世界で他人にアイテムを推薦する状況を考えると、以下のような要素を含むと考えられる。

- 推薦アイテムに対する肯定的な感想
- 被推薦者が興味を持つ内容

このような推薦は、経済活動においても見られる兆候である。例として、本の帯や映画の広告などに付与されている推薦文が挙げられる。

本研究では、Web 上のレビューからこのような推薦文を作成し、ユーザに提示する。また、性能でアイテムの優劣を判断できないアイテムに対してユーザに判断材料を提示することを目的としているため、推薦アイテムとしては「映画」を選択した。アイテムのレビューは、レビュー投稿サイト「みんなのシネマレビュー」<sup>\*4</sup> のレビューを用いる。レビューサイトのレビューはアクセスにかかる時間をなくすために、Web スクレイピングを行いデータベース化した。

### 4.2 提案システムの構成

システムの構成を表す概要図を図 1 に示し、システムの処理の流れを以下に示す。

- (1) ユーザは、既知の映画に対して評価する
- (2) 評価した映画の特徴量を映画データベースから取り出す事でユーザプロフィールを作成する
- (3) ユーザプロフィールと映画データベースを基に、コンテンツフィルタリングによって推薦映画を決定する
- (4) Web 上のレビューとユーザプロフィールを基に推薦映画の推薦文を作成する
- (5) 作成した推薦文と推薦映画をユーザに提示する

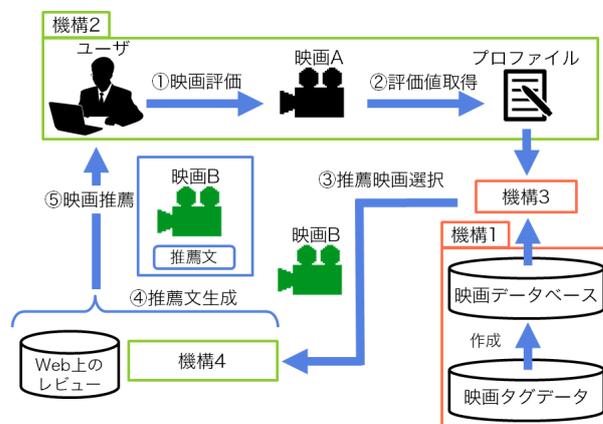


図 1 提案手法の概要図

\*3 NetPromoterScore

\*4 <http://www.jtnews.jp/>

### 4.3 提案システムの機構

提案システム実現のため、以下に述べる機構を実装した。

#### 4.3.1 映画データベース作成機構 (機構 1)

ユーザプロフィール作成の為に各映画の特徴量をデータベース化しておく必要がある。映画の特徴は、movielens\*5で各映画に付与されているタグを用いた。映画  $j$  に付与されているタグ  $i$  の付与回数を  $freq(i, j)$ 、映画  $j$  におけるタグ  $i$  以外のタグ数を  $maxOthers(i, j)$  とすると、映画  $j$  における正規化されたタグ付与回数  $tf(i, j)$  は式 1 で表される。

$$tf(i, j) = \frac{freq(i, j)}{maxOthers(i, j)} \quad (1)$$

$N$  を推薦可能な映画の総数とし、 $n(i)$  をタグ  $i$  が付与されている映画の数とすると、 $i$  の逆文書頻度は式 2 で表される。

$$idf(i) = \log \frac{N}{n(i)} \quad (2)$$

映画  $j$  におけるタグ  $i$  の  $tf \cdot idf$  は式 3 に示す様にこれらの積として定義されている。

$$tf \cdot idf = tf(i, j) \times idf(i) \quad (3)$$

この  $tf \cdot idf$  を映画の特徴量として映画データベースを作成した。

#### 4.3.2 ユーザプロフィール作成機構 (機構 2)

本機構は、ユーザが評価した映画に付与されているタグの重みをユーザプロフィールに追加する。また、既に追加されているタグに関しては重みの加算を行う。これにより、ユーザの嗜好を特徴に対する興味度合いとして表すことができる。

#### 4.3.3 推薦アイテム選択機構 (機構 3)

推薦アイテム選択機構では、内容ベースフィルタリングを用いてアイテム集合からユーザに対して推薦するアイテムを選択する機構である。本機構では、図 2 に示すように、映画データベースとユーザプロフィールを用いてユーザの嗜好と類似度が高い映画を推薦映画として選択する。ユーザプロフィール  $a$  の特徴量ベクトルを  $\mathbf{a}$ 、映画  $b$  の特徴量ベクトルを  $\mathbf{b}$  とすると、 $a, b$  間の類似度は式 4 で示すコサイン類似度で算出した。

$$sim(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}| \times |\mathbf{b}|} \quad (4)$$

#### 4.3.4 推薦文作成機構 (機構 4)

推薦文作成機構は、選択した推薦アイテムに対する推薦文を作成する機構である。本機構の概要図を図 3 に示す。

本研究において、推薦文は「推薦アイテムに対する肯定的な感想」且つ「被推薦者が興味を持つ内容」と仮定した。本機構の処理の流れを以下に示す。

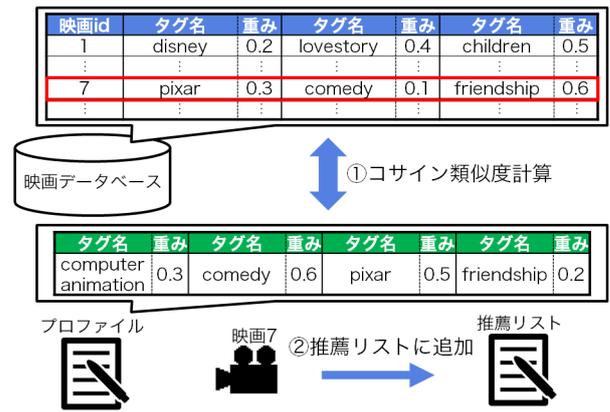


図 2 推薦アイテム選択方法の概要図

- (1) ユーザの興味語の類語を取得する
- (2) レビューデータベース内のレビューから推薦映画のレビューを検索する
- (3) 推薦映画のレビューを、ピリオド、クエスチョンマーク、エクスクラメーションマークで区切り、文単位に分割する
- (4) レビュー文内にユーザの嗜好タグが含まれる文を抽出する
- (5) 4で抽出した文に対して形態素解析を行い、語単位に分割する。形態素解析には MeCab[6] を用いる
- (6) 形態素解析結果から日本語極性辞書を用いて評価表現を検出する
- (7) 評価表現を含む文を推薦文候補とする

レビュー文内にタグが出てこない場合でも、似た表現の単語が出現している場合が考えられる。一例として、「Pixar 映画は子供に大人気で家族で楽しく見えます。」というレビュー文は、ユーザの嗜好タグが「PIXR」であった場合、ユーザへの推薦文候補とはならない。しかし、類似した意味を持つワードが出現しているため、推薦文候補として考慮すべきである。そこで、Weblio 類語辞典\*6を用いる事でこの問題を解決する。Weblio 類語辞典は、単語の類語を検索する Web サービスである。タグの類語を取得する事で、ユーザの嗜好に合うレビュー文を抽出出来る可能性が

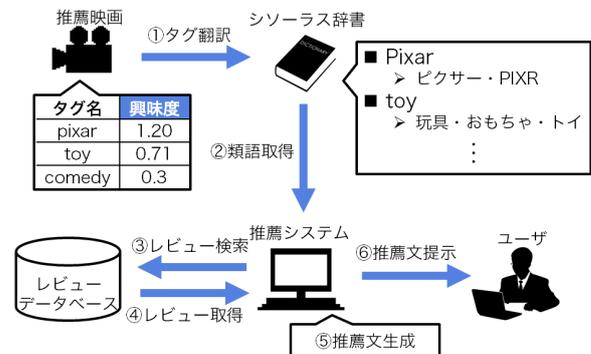


図 3 推薦文作成の概要図

\*5 <http://grouplens.org/datasets/movielens/>

\*6 <http://thesaurus.weblio.jp/>

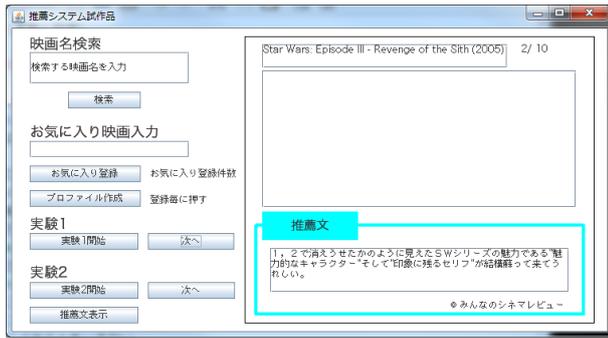


図 4 既知の映画に対する推薦文の付与

向上すると考えられる。

肯定的レビュー文の抽出は、文中に肯定的表現が含まれているかで判断する。肯定的表現の検索には、小林ら [7] の日本語評価極性辞書 (用言編) と東山ら [8] の日本語評価極性辞書 (名詞編) を用いた。日本語評価極性辞書には、名詞及び用言に対して肯定的 ( $P_k = p$ ) か否定的 ( $P_k = n$ ) の評価がなされている。本研究では、レビュー文に含まれる名詞と用言の極性  $P_k$  を基に、式 5 より文極性判別値  $D$  を算出する。  $D > 0$  の場合、文を肯定文とした。

$$D = \begin{cases} D + 1 & (P_k = p) \\ D - 1 & (P_k = n) \\ D & (\text{otherwise}) \end{cases} \quad (5)$$

推薦文候補中で興味語含有数が最大の文を推薦文とする。興味語含有数が同数の文があれば、文極性判別値  $D$  が大きい文を優先する。  $D$  も同値の文があれば、その中からランダムで推薦文を決定する。

## 5. 実験と評価

提案システムの評価を行うためにプロトタイプシステムを作成した。実験は 2 種類の実験を行った。

### 5.1 実験 A

#### 5.1.1 目的と方法

実験 A では、本研究における推薦文の仮定の妥当性を検証することを目的とした。

本学情報工学科の学生 5 名を被験者とし、各被験者は自分の好みの映画を 10 件選択し、ユーザプロフィールの作成を行った。この選択した 10 件の映画に推薦文を付与し、図 4 に示すように 1 件ずつ被験者に提示した。被験者は提示された映画に対する推薦文を読んだ後、以下の質問に対して最高評価を 5 とし、5 段階の評価を行った。

#### A 映画の特徴を表していると思いますか?

5. 非常に表している
4. 表している
3. どちらともいえない
2. 表していない

#### B 映画を賞賛する内容であると思いますか?

5. 非常に賞賛している
4. 賞賛している
3. どちらともいえない
2. 賞賛していない

#### C 映画の推薦文として適切であると思いますか?

5. 非常に適切である
4. 適切である
3. どちらともいえない
2. 適切ではない
1. 全く適切ではない

#### 5.1.2 実験 A の結果と考察

本実験では、全ユーザが登録した映画 50 件中 41 件に推薦文を付与できた。レビューが少なかったため推薦文が提示できなかった映画に関しては結果に含めない事とした。

実験 A における質問の回答を  $\{g_n, g_p\} = \{(1, 2, 3), (4, 5)\}$  のようにグループ分けした。実験 A では、本研究における「ユーザが興味をもつキーワードを含む」且つ「肯定的な文」であれば推薦文となりうるという仮定の妥当性を検証する。そのため、 $A(g_p) \cap B(g_p) = C(g_p)$ ,  $A(g_n) \cup B(g_n) = C(g_n)$  が成り立つかを検証する。

実験 A の結果を表 1, 表 2 に示す。表 1 中の括弧内の値は、 $i = \{p, n\}, j = \{p, n\}$  とすると、 $(A(g_i) \cap B(g_j))/C(g_p)$  を表す。また、表 2 中の括弧内の値は、 $(A(g_i) \cap B(g_j))/C(g_n)$  を表す。

表 1 質問 C における  $g_p$  の割合 [%]

	質問 $B(g_p)$	質問 $B(g_n)$
質問 $A(g_p)$	90(9/10)	0(0/10)
質問 $A(g_n)$	10(1/10)	0(0/10)

表 2 質問 C における  $g_n$  の割合 [%]

	質問 $B(g_p)$	質問 $B(g_n)$
質問 $A(g_p)$	9.7(3/31)	22.6(7/31)
質問 $A(g_n)$	35.5(11/31)	32.3(10/31)

表 1 において  $A(g_p) \cap B(g_p) = C(g_p) = 90[\%]$  であることから、 $A(g_p) \cap B(g_p)$  である場合、推薦文として適切である可能性が高い。一方、表 2 より、 $A(g_n) \cup B(g_n) = C(g_n) = 90.4[\%]$  であることから、 $A(g_n) \cup B(g_n)$  である場合、推薦文として適切である可能性は低い。

また、 $A(g_p) \cap B(g_p)$  における質問 C の回答が  $A(g_n) \cup B(g_n)$  における質問 C の回答よりも有意に高いと言えるかどうかを検証するために、有意水準 5% で Welch の t 検定を行った。その結果、 $A(g_p) \cap B(g_p)$  における質問 C の回答は、 $A(g_n) \cup B(g_n)$  における質問 C の回答より

も有意に高い評価を得るという結果が得られた。以上より、本研究における推薦文の仮定  $A(g_p) \cap B(g_p) = C(g_p)$ ,  $A(g_n) \cup B(g_n) = C(g_n)$  には妥当性があると言える。

## 5.2 実験 B

### 5.2.1 目的と方法

実験 B では推薦文がアイテムの選択要因となりうるかの検証を目的とした。

実験 A で作成したプロフィールを基に、被験者に対して 30 件の映画を推薦した。映画推薦の際に、まず図 5 に示すように推薦文を付与しない状態で映画を推薦した。映画の推薦時には以下の情報を提示した。

- 評価平均点数
- キーワード
- 映画スタッフ

映画を推薦された被験者は、D の質問に対して 5 段階で評価を行った。次に、図 6 に示すように推薦した映画に推薦文を付与し、E, F, G の質問を行った。

D 推薦された映画を見たいと思いますか?

5. 非常に見たい
4. 見たい
3. どちらともいえない
2. 見たくない
1. 全く見たくない

E 提示された推薦文は「映画の推薦文である」と感じましたか?

5. 非常に感じた
4. 感じた
3. どちらともいえない
2. 感じない
1. 全く感じない

F D の質問に対して「非常に見たい・見たい」と回答した方に質問です。

推薦文がある場合とない場合ではどちらの方が映画を見たいと思えますか?

- ON 推薦文がある場合
- OFF 推薦文がない場合



図 5 推薦文を付与しない状態の映画推薦

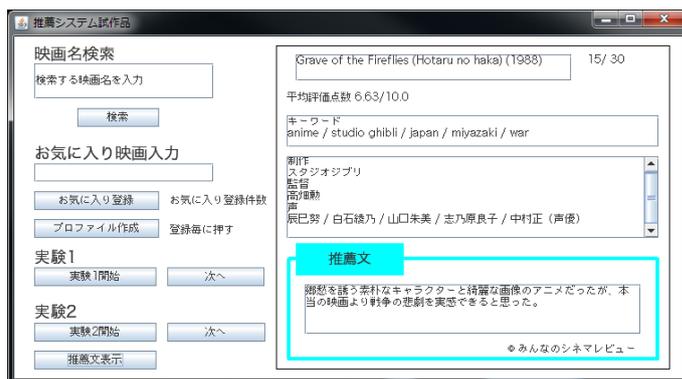


図 6 推薦文を付与した状態の映画推薦

G D の質問に対して「非常に見たくない・見たくない・どちらとも言えない」と回答した方に質問です。

推薦文を読んで、映画を見たいと思えますか?

5. 非常に見たい
4. 見たい
3. どちらともいえない
2. 見たくない
1. 全く見たくない

### 5.2.2 実験 B の結果と考察

本実験では、計 150 件の映画を推薦し、77 件の映画に推薦文を付与できた。レビューが少なかったため推薦文が提示できなかった映画に関しては結果に含めない事とした。また、実験 B における質問の回答を  $\{g_n, g_p\} = \{(1, 2, 3), (4, 5)\}$  のようにグループ分けをした。

質問 E と質問 F に対する回答結果を表 3 に示す。表 3 中の括弧内の値は、 $i = \{p, n\}$ ,  $X = \{ON, OFF\}$  とすると、 $((E(g_i) \cap F(X)) / (D(g_p) \cap E(g_i)))$  を表す。

表 3 質問 E と質問 F の回答結果 [%]

	質問 $E(g_p)$	質問 $E(g_n)$
質問 $F(g_p)$	78.9(26/33)	15.8(3/19)
質問 $F(g_n)$	21.1(7/33)	84.2(16/19)

表 3 において、 $E(g_p) \cap F(ON) = 78.9(\%)$  であることから、推薦文として適切な文章であれば、ユーザが興味を持つ推薦映画に対してさらに興味を喚起しているといえる。一方で、 $E(g_n) \cap F(OFF) = 84.2(\%)$  であることから、推薦文として適切ではない文章は、ユーザの推薦映画に対する興味を低下させる可能性があると考えられるため、不用であるといえる。

質問 E と質問 G に対する回答結果を表 4 に示す。表 4 中の括弧内の値は、 $i = \{p, n\}, j = \{p, n\}$  とすると、 $((E(g_i) \cap G(g_j)) / (D(g_n) \cap E(g_i)))$  を表す。

表 4 において、 $E(g_p) \cap G(g_p) = 62.5(\%)$  であることから、適切な推薦文であれば、推薦文がない状態の推薦に興味があつた映画に対してユーザの興味を喚起させる傾向があるといえる。

表 4 質問 E と質問 G の回答結果 [%]

	質問 E( $g_p$ )	質問 E( $g_n$ )
質問 G( $g_p$ )	62.5(5/8)	0.0(0/17)
質問 G( $g_n$ )	37.5(3/8)	100.0(17/17)

本研究で作成した推薦システムの精度 (accuracy) を質問 D より算出し, 表 5 に示す. レビュー文付きで推薦できた映画の総数を  $M$  とし, 式 6 を用いて算出した.

$$accuracy = \frac{D(g_p)}{M} \quad (6)$$

表 5 本研究で作成した推薦システムの精度 [%]

	推薦文付与前	推薦文付与後
推薦精度	67.5(52/77)	77.9(60/77)

表 5 より, 推薦文付与前に比べて推薦文付与後の精度が 10.4% 向上していることが分かる. これは, 推薦文を付与する事でユーザの興味が無かった映画に対して興味を喚起した結果であるといえる.

実験 B の結果より, 推薦文の提示はユーザにとって推薦アイテムの選択要因となりうると思われる. しかし, 表 4 は標本数が少ないためさらなる実験を重ねる必要がある. また, 150 件中 73 件の映画に推薦文を付与出来なかったため, データベースからシステムの見直しを行う必要があると思われる.

## 6. まとめ

内容に即した推薦文を提示するシステムを提案し, 評価するためのプロトタイプシステムを作成した. 本研究における推薦文の仮定の妥当性を検証する実験 A と, 推薦文がアイテム選択要因として有用かを検証する実験 B を行った.

実験 A においては仮定に対して十分な妥当性を示す事ができた. 実験 B においては, 推薦文がアイテム選択要因となり得る可能性があることを示唆する以下の二点の傾向を確認した.

- 推薦文の提示はユーザの興味を喚起する
- 推薦文を提示しない状態で興味が無かった映画に対して, 推薦文を提示する事で興味を喚起する

今後の予定について述べる.

### 6.1 否定語による単語の極性反転への対応

現在, 否定語によって単語の極性が反転してしまう現象を確認している. これにより, 本来は否定文であるレビュー文を肯定文として認識してしまう. この問題を解決するために, レビュー文の係り受け解析を行い, 単語に対して意味を反転させる語が付随している場合, 単語の極性を反転させる機能を追加する.

### 6.2 新たな特徴データベース作成方法の検討

本研究では, 映画データベースの作成に movielens のタグデータを用いた. しかし, ユーザにアイテムのタグを付与させて特徴データベースを作成することは膨大な時間がかかる.

そこで, レビューから映画の特徴付けを行うことで, データベースの作成を自動化する.

### 6.3 協調フィルタリングにおける推薦文作成方法の考案

本研究では, 推薦文が推薦アイテムの推薦理由として機能し, ユーザのアイテム選択要因となると仮定し, 内容ベースフィルタリングを採用していた. しかし, 現在導入されている推薦システムは協調フィルタリングが主流である. 提案システムの汎用性を高めるために協調フィルタリングにおける推薦文作成方法を考案する必要がある. 協調フィルタリングは, アイテムの特徴データベースが不要であるという点が利点である. そこで, 今後は前述したようにデータベースの作成を自動化することで, 協調フィルタリングの利点を損なわずに推薦文作成システムの導入を試みる.

### 参考文献

- [1] Renjie Zhou, Samamon Khemmarat, Lixin Gao, "The Impact of YouTube Recommendation System on Video Views", In Proc. of the IMC '10 Proceedings of the 10th ACM SIGCOMM conference on Internet measurement, pp.404-410, (2010).
- [2] Dietmar Jannach and Markus Zanker and Alexander Felferning and Gerhard Friedrich, "情報推薦システム入門", 共立出版.
- [3] 折原 レオナルド賢, 橋山 智訓, 田野 俊一, "推薦理由を提示する情報推薦システム", 人工知能学会全国大会, Vol.29, (2015).
- [4] Yan Lin, "嗜在日留学生のアルバイト活動における動機づけと仕事満足感", 日本心理学会, vol.78, (2014).
- [5] 林貴宏, 尾内理紀夫, "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).
- [6] Taku Kudo et al., "Applying Conditional Random Fields to Japanese Morphological Analysis", proceedings of the 2004. Conference on Empirical Methods in Natural Language Processing (EMNLP-2004), pp.230-237 (2004).
- [7] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一, "意見抽出のための評価表現の収集", 自然言語処理, vol.12, No.2, pp.203-220, (2005).
- [8] 東山昌彦, 乾健太郎, 松本裕治, "述語の選択嗜好性に着目した名詞評価極性の獲得", 言語処理学会発表論文集, vol.14, pp.584-587, (2008).