

# 利用者の潜在的嗜好を予測する協調フィルタリングの検討

Ta Son Tung<sup>†</sup> 奥 健太<sup>†</sup> 服部 文夫<sup>†</sup>

<sup>†</sup>立命館大学情報理工学部 〒525-8577 滋賀県草津市野路東1丁目1-1

E-mail: †{cc0000pi@ed, oku@fc, fhattori@is}.ritsumeai.ac.jp

**あらまし** 利用者の嗜好への適合性以外に、いかに利用者が知らないアイテムを推薦するかといった新規性や発見性を考慮した情報推薦が重要視されてきている。しかしながら、利用者が知らないジャンルについては利用者の嗜好に関するデータが乏しく嗜好を正確に予測することが困難である。本研究では、ジャンル間の関係性を考慮し、既知ジャンルにおける嗜好から未知ジャンルにおける潜在的な嗜好を協調フィルタリングにより予測する方式を提案する。

**キーワード** 情報推薦, 潜在的嗜好, 新規性

## A Method to predict Potential Preferences of Users using Collaborative Filtering

Ta Son Tung<sup>†</sup> Kenta OKU<sup>†</sup> and Fumio HATTORI<sup>†</sup>

<sup>†</sup>Department of Information and Computer Science, Ritsumeikan University

1-1-1 Noji-higashi, Kusatsu-shi, Shiga, 525-8577 JAPAN

E-mail: †{cc0000pi@ed, oku@fc, fhattori@is}.ritsumeai.ac.jp

### 1. はじめに

インターネットの普及とともに、日々膨大な情報が生み出されるようになり、利用者は自分の興味がある情報や商品(以降アイテムと呼ぶ)を探すことが困難になった。このような問題を解決するために、情報推薦技術に関する研究が行われてきた[1]。情報推薦技術は利用者(以降ユーザと呼ぶ)の購入履歴や嗜好・デモグラフィック情報などに基づいてユーザの嗜好に合ったアイテムを提供する技術である。

情報推薦技術には内容に基づくフィルタリングと協調フィルタリングという2つの主な方式がある[2]。その中では特に協調フィルタリングがよく使われている。成功している情報推薦システムの代表例としてはAmazon.comが挙げられる[3]。既存の推薦システムではユーザの嗜好に忠実なアイテムを推薦するように、推薦に対する適合性を重視したものが多く。しかし、嗜好への適合性が高くても、必ずしも推薦されたアイテムがユーザにとって有用であるとは限らない。例えば、ユーザが既によく知っているアイテムはユーザの嗜好に合ったものであったとしても推薦する価値は高くない[4][5]。推薦システムには、いかにユーザにとって未知であり、かつ有用であるアイテムを推薦するか

が求められる。

本研究では、推薦対象とするアイテムをいくつかのジャンルに分けたとき、ユーザが既に知っているジャンル(既知ジャンル)と、知らないジャンル(未知ジャンル)に着目する。例えば、映画を推薦対象アイテムとした場合、「アクション」や「ホラー」、「ラブストーリー」などがジャンルにあたる。「アクション」をよく観るユーザにとっては「アクション」がそのユーザにとっての既知ジャンルとなり、「ホラー」をほとんどもしくはまったく観ないユーザにとっては「ホラー」がそのユーザにとっての未知ジャンルとなる。このとき、未知ジャンルについては、(a)そのユーザにとっては本当に興味がない、(b)潜在的には興味があるがユーザ自身がそれに気付いていない、ということが考えられる。もし(b)の場合であれば、推薦システムにより推薦することでユーザにとって新たな気付きを促すことができる。本研究では、ユーザの未知ジャンルにおける嗜好を潜在的嗜好と定義し、この潜在的嗜好を予測する推薦方式を確立することを目的とする。

しかしながら、対象ユーザにとっての未知ジャンルにおいては、そもそもそのジャンルに関する嗜好データ(アイテムに対する評価値データ)がほとんどなく、

従来の推薦方式による嗜好の予測は難しい。そこで、本研究では、既知ジャンルにおける嗜好データから未知ジャンルにおける潜在的嗜好を予測する協調フィルタリング方式を提案する。まず、着目しているジャンルにおける嗜好データの量から、そのジャンルが対象ユーザにとって既知であるか未知であるか判断する。つづいて、後述する方法でジャンル間の関連度を求め、嗜好を抽出しようとしている未知ジャンルと関連する既知ジャンルにおいて、対象ユーザと嗜好が類似する他ユーザを探す。そして、その類似ユーザの嗜好データを参照することで、未知ジャンルにおけるユーザの嗜好を予測する。

本論文の構成は以下の通りである。2章では、関連研究について説明する。3章では、提案手法について説明する。4章では実験および考察について述べる。5章で本論文をまとめ、今後の課題について述べる。

## 2. 関連研究

### 2.1. 情報推薦の基本方式

情報推薦に関してはこれまでに数多くの研究開発が行われてきた。情報推薦技術は大きく次の3種類に分類される[2][6][7]。

内容に基づくフィルタリング: ユーザの購入履歴を用いて、購入したアイテムの内容と推薦対象アイテムの内容を比較し、類似度を計算する。類似度の高い順にアイテムを並べた推薦リストを作成する。この手法ではユーザの購入履歴と関連性の低いアイテムは推薦されにくい。

協調フィルタリング: ユーザベース協調フィルタリングとアイテムベース協調フィルタリングの2種類がある。ユーザベース協調フィルタリングはユーザ間の類似度を計算し、嗜好が類似したユーザが好むアイテムを対象ユーザに推薦する。代表的なシステムにはMovieLens[8]がある。一方、アイテムベース協調フィルタリングはユーザの評価に基づいてアイテム間の類似度を計算し、対象ユーザがあるアイテムに関する情報を閲覧しているとき、このアイテムに類似するアイテムを推薦する。商用化されている代表的なシステムにはAmazon.com[3]がある。

ハイブリッド手法: 内容に基づくフィルタリングと協調フィルタリングを組み合わせた手法である。組み合わせる方法としては、Mixed, Weighted, Switching, Cascade, Meta-level, Feature Combination, Feature augmentationなどが提案されている[2][7]。代表的なシステムとしてはLibra, EntreeC, Fabがある。

### 2.2. 適合性以外の評価指標を考慮する推薦手法

既存推薦システムではユーザの嗜好に忠実なアイテムを推薦するように、推薦に対する適合性を重視し

たものが多い。しかし、近年、適合性以外にも、ユーザ満足度に影響を及ぼす推薦の指標に関して多く研究が行われている[4][5][9][10]。これらの文献において、意外性や新規性[4][5]、多様性[9]、発見性[10]などが提案されているが、これらの定義は各研究によって様々である。

Herlockerら[4]は、意外性(serendipity)と新規性(novelty)がユーザ満足度に影響を及ぼすことを示した。意外性のある推薦とはユーザに驚きを与えながら、好みのアイテムを推薦することである。一方、新規性(novelty)のある推薦とは、推薦されたアイテムがユーザの知らない好みのアイテムを推薦することである。したがって、意外性がある推薦は新規性もあるといえる。新規性や意外性を向上させるため、Herlockerらはユーザにとって自明であるようなアイテムリストを作成し、推薦リストから自明なアイテムを削除するという方法があると述べている。しかし、自明なアイテムはユーザによって異なるという欠点があるとも述べている。そして、人気アイテムよりも、注目されるユーザの嗜好に合ったアイテムのほうを推薦することによって、意外性を向上させる可能性があるとも述べている。

Zieglerら[9]は多様性(diversity)がユーザ満足度に影響を及ぼすことを示した。Zieglerらは元々定義されているアイテムの分類を用いて(具体的にはAmazonのTaxonomy)推薦リスト内のトピックを多様化することで、ユーザの満足度を向上させることを目的としている。しかし、実験の結果によって多様性は向上するが、適合性は低下することが示されている。

村上ら[5]は意外性を考慮した推薦システムを提案し、テレビ番組の推薦に適用した例を報告している。ここで、村上らの研究ではユーザのアンケートで過去に聞いたことがないが興味を持った番組、あるいは、番組を知ってはいたが観たことがなく興味のある番組は意外性のある番組だと判断している。村上らの研究ではユーザが習慣的に選択するコンテンツを予測する習慣モデルとユーザが好むコンテンツを予測する嗜好モデルの2種類の予測モデルを導入している。そして嗜好モデルによる予測結果と習慣モデルによる予測結果との差異に基づいてコンテンツの意外性を推定する。

土方ら[10]はユーザから明示的に与えられる既知・未知というプロフィール情報とユーザの嗜好のプロファイルを結合することによって発見性を向上させる手法を提案している。しかしながら、ユーザから既知・未知の情報を得るにはユーザの労力が大きくなる。

## 3. 提案手法

本研究では、ユーザにとって既知のジャンル(以降、

既知ジャンル)における嗜好から、未知のジャンル(以降、未知ジャンル)における潜在的な嗜好を予測する協調フィルタリング方式を提案する。ここで、対象ユーザー  $u$  とジャンル  $k$  を考える。ジャンル  $k$  中のアイテムに対し、対象ユーザー  $u$  による評価データ数が十分に大きい場合、ジャンル  $k$  を対象ユーザー  $u$  にとって既知ジャンルとよぶ。一方で、対象ユーザー  $u$  による評価データがない、もしくは十分小さい場合、ジャンル  $k$  を対象ユーザー  $u$  にとって未知ジャンルとよぶ。

提案手法では、まず事前にジャンル間の関連度および各ジャンルにおけるユーザ間の類似度を算出しておく。その上で、ジャンル間関連度およびユーザ間類似度を用いた協調フィルタリングにより、未知ジャンルにおける対象ユーザの潜在的嗜好を予測する。以下、3.1 節では既知・未知ジャンルの判定方法について述べる。3.2 節ではジャンル間関連度の算出方法、3.3 節ではユーザ間類似度の算出方法についてそれぞれ述べ、3.4 節において潜在的嗜好を予測する協調フィルタリング方式について説明する。

### 3.1. 既知・未知ジャンルの判定

ジャンル  $k$  に着目したとき、ジャンル  $k$  に含まれるアイテム集合を  $I_k$  とし、 $I_k$  内のアイテムの総数を  $|I_k|$  とする。ここで、アイテム集合  $I_k$  に対する対象ユーザー  $u$  による評価データ数を  $n$  とする。パラメタ  $\alpha$  ( $0 \leq \alpha \leq 1$ ) を考えたとき、 $n \geq \alpha |I_k|$  の場合、ジャンル  $k$  は対象ユーザー  $u$  にとって既知ジャンルであるとし、 $\text{known}(u, k)=1$  で表す。一方で、 $n < \alpha |I_k|$  の場合、ジャンル  $k$  は対象ユーザー  $u$  にとって未知ジャンルであるとし、 $\text{known}(u, k)=0$  で表す。

ここで、パラメタ  $\alpha$  は実験的に求める必要があるが、本論文では、対象データセットに対する評価データの密度を  $\alpha$  とする。すなわち、 $\alpha$  は次式により求められる

$$\alpha = \frac{N}{|U| \times |I|} \quad (1)$$

ここで、 $N$  は評価データ数、 $|U|$  はデータセットの全ユーザ数、 $|I|$  は全アイテム数である。

### 3.2. ジャンル間類似度の算出

本研究では、二つのジャンルに対し、両ジャンル共に既知となっているユーザが多いほど、その二つのジャンルが関連していると仮定し、この仮定に基づきジャンル間の関連度を算出する。

ジャンル  $k$  が既知であるユーザ集合を  $U_k$  とすると、そのユーザ数  $|U_k|$  は、

$$|U_k| = \sum_i \text{known}(u_i, k) \quad (2)$$

となる。同様に、ジャンル  $l$  が既知であるユーザ集合

を  $U_l$ 、ジャンル  $k$  およびジャンル  $l$  の両方が既知であるユーザ集合を  $U_{k \cap l}$  とする。このとき、ジャンル  $l$  に対するジャンル  $k$  の関連度は次式より求められる。

$$\text{rel}(k, l) = \frac{|U_{k \cap l}|}{|U_k|} \quad (3)$$

ここで、ジャンル  $l$  に対するジャンル  $k$  の関連度  $\text{rel}(k, l)$  とジャンル  $k$  に対するジャンル  $l$  の関連度  $\text{rel}(l, k)$  は異なる。つまり、

$$\text{rel}(k, l) = \frac{|U_{k \cap l}|}{|U_k|} \neq \text{rel}(l, k) = \frac{|U_{k \cap l}|}{|U_l|}$$

となる。例えば、"action"の映画が既知であるユーザが100人であり、"horror"の映画が既知であるユーザが25人、その両方が既知であるユーザが20人であるとする。このとき、"horror"に対する"action"の関連度  $\text{rel}(\text{action}, \text{horror})$ 、"action"に対する"horror"の関連度  $\text{rel}(\text{horror}, \text{action})$  は、それぞれ、次式のようになる。

$$\text{rel}(\text{action}, \text{horror}) = \frac{20}{100} = 0.2 \quad \text{rel}(\text{horror}, \text{action}) = \frac{20}{25} = 0.8$$

つまり、"horror"が既知であるユーザのうち80%のユーザが"action"も既知であり、関連度が高いことを意味する。一方で、"action"が既知であるユーザのうち"horror"も既知であるユーザは20%のみであり、関連度は低いということになる。

### 3.3. ジャンル内ユーザ間類似度の算出

ユーザ間の類似度算出方式としてはいくつかのものが提案されているが、Resnickら(1994)[1]が用いた方法が有名である。従来の協調フィルタリングではユーザのすべての評価データを用いてユーザ間類似度を算出するものが多い。しかしながら、ジャンルが異なれば嗜好も異なる可能性があると考えられるため、本研究ではジャンルごとにユーザ間類似度を算出する。特に、対象ユーザーにとっての既知ジャンルに着目し、既知ジャンル内のアイテムに対する評価データからユーザ間類似度を算出する。

対象ユーザーにとっての既知ジャンル  $k$  が既知である他ユーザ集合を  $U_k = \{u_1, u_2, \dots, u_n\}$ 、ジャンル  $k$  内のアイテム集合を  $I_k = \{i_1, i_2, \dots, i_m\}$  とする。また、ユーザ  $u_i$  のアイテム  $i_j$  に対する評価値を  $r_{ij}$  とする。このとき既知ジャンル  $k$  におけるユーザ  $u_p$  とユーザ  $u_q$  の類似度は次式より求められる。

$$\text{sim}_k(u_p, u_q) = \frac{\sum_j (r_{pj} - \bar{r}_p)(r_{qj} - \bar{r}_q)}{\sqrt{\sum_j (r_{pj} - \bar{r}_p)^2} \sqrt{\sum_j (r_{qj} - \bar{r}_q)^2}} \quad (4)$$

ここで、 $\bar{r}_p = \left(\frac{1}{|I_p|}\right) \sum_{i \in I_p} r_{pi}$  ;  $I_p = \{i \in I | r_{pi} \neq \phi\}$  である。

### 3.4. 未知ジャンルにおける潜在的嗜好の予測

3.2 節および 3.3 節で算出されたジャンル間関連度お

よびジャンル内ユーザ間類似度を用いて、未知ジャンルにおける対象ユーザの潜在的嗜好を予測する。

未知ジャンル内のアイテムに対する嗜好については、従来の協調フィルタリングのとおり、嗜好が類似するユーザを重視して予測する。さらに本研究では、嗜好予測対象である未知ジャンルとのジャンル間関連度が高い既知ジャンルにおいて嗜好が類似するユーザを重視する。この考えを取り入れると、対象ユーザ  $u_p$  の未知ジャンル  $x$  内のアイテム  $i_y$  に対する嗜好の予測値  $p(x,y)$  は以下の手順により求められる。

(i) 未知ジャンル  $x$  に対する関連度が最も高い  $m$  個のジャンル集合  $G=\{g_1, g_2, \dots, g_m\}$  を選ぶ。

(ii) 各ジャンル  $g_i$  ( $i=1, 2, \dots, m$ ) において、対象ユーザ  $u_p$  との類似度が最も高い  $j$  人のユーザ集合  $V=\{v_1, v_2, \dots, v_j\}$  を選ぶ。

(iii) 選び出されたユーザの嗜好データを参照し、ジャンル内各アイテムの評価値を予測する

上記の手順を式で表すと次式のようになる。

$$p(x,y) = \bar{r}_p + \frac{\sum_{k \in G} \sum_{u_q \in V} \text{rel}(x,k) \text{sim}_k(u_p, v_q) (r_{qy} - \bar{r}_q)}{\sum_{k \in G} \sum_{v_q \in V} |\text{rel}(x,k)| |\text{sim}_k(u_p, v_q)|} \quad (5)$$

ここで、 $\bar{r}_p$  は以下の式で求める

$$\bar{r}_p = (1/|I_p|) \sum_{i \in I_p} r_{pi} \quad I_p = \{i \in I | r_{pi} \neq \phi\}$$

最終的に、予測評価値  $p(x,y)$  に基づいて上位  $N$  個の推薦リスト  $L=\{1, 2, \dots, N\}$  が求められる。

#### 4. 実験及び考察

提案手法の有用性を評価するために実験を行った。本実験では、ジャンル間の関連度がユーザの潜在的嗜好の予測にどの程度寄与するか検証する。実験にはデータセットとして GroupLens から公開されている MovieLens Data Sets 100k [1] を利用する。本データセットの概要は表 1 のとおりである。

実験の手順は以下のとおりとなる。

(i) 3.1 節で述べたように、事前に各ジャンルに対し、既知・未知のユーザ集合を取得する。ここで、 $\alpha$  は評価データの密度に相当するため、 $\alpha = 0.063$  となる。各ジャンルの映画総数および既知・未知の閾値 ( $\alpha |I_k|$ ) を表 2 に示す。また、その閾値を超える評価データを持つユーザ、すなわち既知ユーザの数を併せて表 2 に示す。

(ii) 表 2 に示したように、MovieLens Data Sets には 18 のジャンルが含まれる。それぞれのジャンル間の関連度を算出する。表 3 に各ジャンル間の関連度の算出結果を示す。表 3 の行列が各ジャンルを表し、交わる

表 1 MovieLens Data Sets の概要

評価アイテム (映画データ)数	1680(本)
評価ユーザ数	943(人)
評価データ数	99990(件)
ジャンル数	18(件)

表 2 各ジャンルの既知ユーザ数

ジャンル Id	ジャンル 名	映画 総数	$\alpha  I_k $	既知 ユーザ数
0	Action	251	16	491
1	Adventure	135	9	492
2	Animation	42	3	362
3	Children	122	8	283
4	Comedy	505	32	324
5	Crime	109	7	413
6	Documentary	50	4	86
7	Drama	725	46	308
8	Fantasy	22	2	294
9	Film-Noir	24	2	355
10	Horror	92	6	269
11	Musical	56	4	386
12	Mystery	61	4	544
13	Romance	247	16	427
14	Sci-Fi	101	7	549
15	Thriller	251	16	464
16	War	71	5	602
17	Western	27	2	345

部分の数値が対象ジャンル間の関連度を示す。関連度の強さが分かりやすいように、表 3 では対応するセルをグレースケール(薄い⇒関連度低、濃い⇒関連度高)で表している。なお、3.1 節で述べたとおり、ジャンル間の関連度は非対称となる

本実験では、表 4 に示す延べ 12 個のジャンルを未知ジャンルとして扱う。各ジャンルに対して、関連度が高いジャンルおよび低いジャンルを各 1 ジャンルずつ選択する。

さらに、表 4 に示す各三つのジャンルがすべて既知であるユーザを選択する。選択されたユーザ数を表 4 に示す。また、あらかじめ関連度が高いジャンルおよび低いジャンルそれぞれにおける評価データを用いてユーザ間の類似度を算出しておく。本実験では、これら関連度が高いジャンルおよび低いジャンルにおけるユーザの評価データから未知ジャンルにおけるユーザの潜在的嗜好を予測する。

表 3 : ジャンル間関連

id	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0	1	0.92	0.65	0.54	0.63	0.72	0.15	0.59	0.57	0.57	0.52	0.70	0.75	0.76	0.96	0.86	0.93	0.63
1	0.92	1	0.66	0.56	0.62	0.70	0.15	0.59	0.56	0.55	0.5	0.72	0.72	0.76	0.95	0.81	0.92	0.63
2	0.88	0.90	1	0.73	0.72	0.70	0.17	0.64	0.66	0.58	0.55	0.84	0.75	0.81	0.92	0.82	0.93	0.68
3	0.94	0.97	0.93	1	0.84	0.77	0.19	0.73	0.81	0.64	0.66	0.97	0.81	0.92	0.96	0.89	0.96	0.80
4	0.95	0.94	0.81	0.73	1	0.85	0.22	0.79	0.74	0.66	0.66	0.89	0.85	0.97	0.96	0.92	0.98	0.77
5	0.86	0.83	0.61	0.53	0.67	1	0.19	0.66	0.56	0.67	0.57	0.69	0.85	0.80	0.89	0.89	0.93	0.66
6	0.88	0.88	0.72	0.63	0.84	0.91	1	0.88	0.65	0.87	0.60	0.87	0.95	0.93	0.91	0.90	0.97	0.83
7	0.94	0.94	0.75	0.67	0.83	0.89	0.25	1	0.68	0.75	0.66	0.87	0.93	0.97	0.95	0.93	0.99	0.81
8	0.95	0.94	0.81	0.78	0.82	0.79	0.19	0.71	1	0.63	0.65	0.86	0.81	0.88	0.95	0.90	0.94	0.73
9	0.79	0.76	0.59	0.51	0.61	0.77	0.21	0.65	0.52	1	0.54	0.69	0.91	0.74	0.83	0.80	0.90	0.65
10	0.95	0.91	0.74	0.70	0.80	0.87	0.19	0.75	0.71	0.71	1	0.83	0.87	0.87	0.96	0.96	0.93	0.78
11	0.89	0.91	0.79	0.71	0.75	0.74	0.19	0.69	0.67	0.64	0.58	1	0.79	0.89	0.92	0.83	0.95	0.75
12	0.68	0.65	0.50	0.42	0.51	0.64	0.15	0.53	0.44	0.59	0.43	0.56	1	0.64	0.71	0.72	0.79	0.52
13	0.87	0.87	0.68	0.61	0.74	0.77	0.19	0.70	0.60	0.62	0.55	0.80	0.81	1	0.89	0.85	0.97	0.70
14	0.85	0.85	0.61	0.49	0.56	0.67	0.14	0.53	0.51	0.53	0.47	0.65	0.71	0.70	1	0.76	0.90	0.57
15	0.91	0.86	0.64	0.54	0.64	0.79	0.17	0.62	0.57	0.61	0.55	0.69	0.84	0.78	0.89	1	0.91	0.62
16	0.76	0.75	0.56	0.45	0.52	0.64	0.14	0.51	0.46	0.53	0.41	0.61	0.71	0.69	0.82	0.70	1	0.56
17	0.89	0.90	0.71	0.66	0.73	0.79	0.21	0.73	0.63	0.67	0.61	0.83	0.82	0.87	0.91	0.83	0.98	1

(iii) 本実験においては、あるユーザを対象としたとき、表 4 に示したように三つの既知ジャンルのうち一つを未知ジャンルとして扱う。ここで、その未知ジャンルと関連度の高いジャンルおよび低いジャンルそれぞれにおけるユーザ間類似度を算出する。そして、3.4 節で述べた方法により未知ジャンルにおける対象ユーザの潜在的嗜好を予測する。結果、その予測値と実際に与えられている評価値から平均絶対誤差 (MAE) [11]を求める。

実験の結果を図 1 に示す。図 1 はジャンル間関連度と MAE の関係を示している。

表 4 : 実験対象ジャンルと既知ユーザ数および実験により得られた MAE

未知ジャンル	関連度が高いジャンル			関連度が低いジャンル			既知ユーザ数
	Id	Rel	MAE	Id	Rel	MAE	
C <sub>0</sub>	C <sub>14</sub>	0.96	0.58	C <sub>9</sub>	0.57	0.87	273
C <sub>0</sub>	C <sub>1</sub>	0.92	0.55	C <sub>2</sub>	0.64	0.90	309
C <sub>5</sub>	C <sub>16</sub>	0.93	0.34	C <sub>10</sub>	0.57	1.44	275
C <sub>0</sub>	C <sub>16</sub>	0.93	0.77	C <sub>9</sub>	0.57	1.39	228
C <sub>17</sub>	C <sub>16</sub>	0.98	0.39	C <sub>3</sub>	0.66	0.82	225
C <sub>5</sub>	C <sub>14</sub>	0.89	0.36	C <sub>4</sub>	0.67	1.60	269
C <sub>12</sub>	C <sub>15</sub>	0.72	0.72	C <sub>2</sub>	0.50	1.20	253
C <sub>12</sub>	C <sub>15</sub>	0.72	0.47	C <sub>2</sub>	0.50	1.18	52
C <sub>6</sub>	C <sub>12</sub>	0.95	0.27	C <sub>10</sub>	0.60	1.11	179
C <sub>6</sub>	C <sub>12</sub>	0.95	0.39	C <sub>10</sub>	0.60	1.18	273
C <sub>8</sub>	C <sub>1</sub>	0.94	0.51	C <sub>9</sub>	0.63	0.77	309
C <sub>8</sub>	C <sub>1</sub>	0.94	0.12	C <sub>9</sub>	0.63	0.52	275

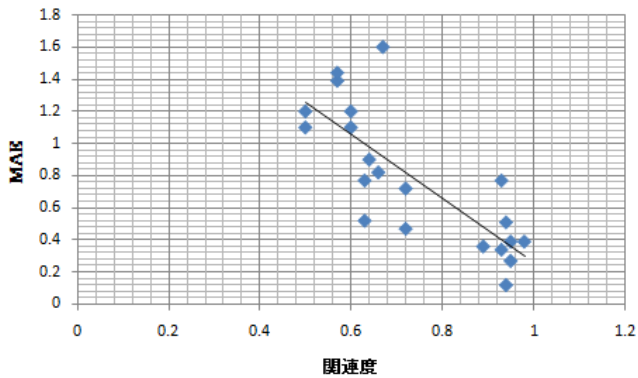


図 1 : 関連度と平均絶対誤差の関連

図 1 によって、関連度が高ければ、高いほど MAE が低くなるということが分かる。すなわち、ジャンル間の関連度が未知ジャンルにおける潜在的嗜好への予測精度に影響を及ぼすことを明らかにした。

## 5. まとめと今後の課題

本研究では、ジャンル間の関連度を考慮し、既知ジャンルにおけるユーザの嗜好から未知ジャンルにおける潜在的嗜好を協調フィルタリングにより予測する方式を提案した。提案手法では、まず事前にジャンル間の関連度および各ジャンルにおけるユーザ間の類似度を算出しておく。その上で、ジャンル間関連度およびユーザ間類似度を用いた協調フィルタリングにより、未知ジャンルにおける対象ユーザの潜在的嗜好を予測する。MovieLens データセットにより、本提案手法の有用性を確認した。

今回の実験では、既知・未知ジャンルとして一つずつ取り出したが、今後は複数のジャンル多くのを考慮に入れ、提案手法に関する詳細な実験を行う。また、映画だけでなく、音楽や本などのデータセットを対象にした評価実験を行う。

## 参 考 文 献

- [1] Resnick, P., N. Iakovou, M. Sushak, P. Bergstrom, and J. Riedl. GroupLens: An open architecture for collaborative filtering of netnews, In Proceedings of the 1994 Computer Supported Cooperative Work Conference, 1994
- [2] Balabanovic, M. and Y. Shoham. Fab: Content-based, collaborative recommendation. Communications of the ACM, 40(3):66-72, 1997
- [3] Linden, G., B. Smith, and J. York. Amazon.com Recommendations: Item-to-Item Collaborative Filtering. IEEE Internet Computing, Jan.-Feb. 2003.
- [4] Herlocker, J., Konstan, J., Terveen, L. and Riedl, J.: Evaluating Collaborative Filtering Recommender System, J. of ACM Transactions on Information Systems, Vol.22, No.1, pp.5-53, 2004.
- [5] 村上智子, 森紘一郎, 折原良平: 推薦の意外性向上のための手法とその評価, 人工知能学会論文誌

24 卷 5 号 G 2009 年.

- [6] Adomavicius, G., Tuzhilin, A.: Toward the Next Generation of RS: A survey of the State-of-the-Art and Possible Extensions, 2005.
- [7] Bruke, R.: Hybrid Recommender Systems: Survey and Experiments, (pre2000) 2002.
- [8] Miller, B. N., I. Albert, S. K. Lam, J. A. Konstan, and J. Riedl. MovieLens Unplugged: Experiences with an Occasionally Connected Recommender System, In Proceedings of the International Conference on Intelligent User Interfaces, Miami, Florida, 2003.
- [9] Ziegler, C. N., Mcnee, S. M., Konstan J.A. and Lausen G.: Improving Recommendation Lists Through Topic Diversification, In pro.of World Wide Web Conference, pp.22-32, 2005.
- [10] Hijikata, Y., Shimizu, T., Nishida, S.: Discovery-oriented Collaborative Filtering for Improving User Satification, In pro. of the 14<sup>th</sup> ACM International Conference on Intelligent User Inter-faces (ACM IUI 2009), pp.67-76, 2009.
- [11] Herlocker, J., Konstan, J., Borchers, A., and Riedl, J., An algorithmic framework for performing collaborative filtering, In Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (Berkeley, CA, USA, 1999), pp. 230-237.