

修士論文

2005（平成 17）年度

論文題目

活動利用者の入力コスト削減を
目的とした推薦アルゴリズムの提案

慶應義塾大学大学院理工学研究科
開放環境科学専攻

学籍番号 80420489 森 崇志

指導教員 教授 櫻井 彰人

慶應義塾大学大学院理工学研究科

2006 年 3 月

謝辞

本論文を作成するにあたり、数々の助言を下された慶応大学大学院理工学研究科開放環境科学専攻の櫻井彰人教授に深く感謝申し上げます。また、中間発表の場で適切な助言を下された山口高平教授、飯島正先生、本研究の手法を実装するに当たり様々な助言を下された篠沢佳久先生に感謝申し上げます。さらに、管理工学科計算機系の諸先輩方、同輩・後輩の皆様、および櫻井・山口研究室の皆様にご心より感謝致します。加えて、被験者として本実験の検証に協力して下さった皆様にごこの場を借りて深く感謝いたします。最後に終始支援してくれた両親に深く感謝いたします。

目次

第1章	序論	1
1.1	背景	2
1.2	本論文の構成	3
第2章	協調フィルタリング	4
2.1	データ収集	5
2.2	類似ユーザー探索	6
2.3	推薦アイテム推定	7
第3章	既存システムの問題点と本研究の目的	8
3.1	本研究の着眼点	9
3.2	既存システムの現状分析と問題点	9
3.2.1	新規ユーザー問題 (start-up 処理)	9
3.2.2	ユーザーの評価手法 (SD法、順序応答)	11
3.3	本研究の目的	12
3.3.1	活動利用者の入力コストの削減	12
3.3.2	「見たい」映画を推薦するシステム	13
第4章	本研究の経緯	14
4.1	本研究で用いる学習データ	15
4.2	本研究当初のアプローチ	15
4.3	数量化理論	17
4.3.1	数量化Ⅲ類	17
4.3.2	数量化Ⅳ類	19
4.3.3	特異値分解	20
4.4	映画情報を視覚化した前例	21
4.5	アプローチの検証	21
第5章	提案手法	26
5.1	データ収集	28
5.2	活動利用者の嗜好範囲の特定	29
5.3	推薦アイテム決定	32
第6章	予備実験	37

6. 1	予備実験の条件	38
6. 2	予備実験の内容	38
6. 2. 1	映画の選択	38
6. 2. 2	一対比較	39
6. 2. 3	推薦映画の評価	40
6. 3	実験条件の決定	41
6. 3. 1	解析手法	41
6. 3. 2	次元数 n	43
6. 3. 3	一対比較数 k	43
第7章 本実験とその評価・考察		47
7. 1	実験条件	48
7. 2	実験内容	48
7. 3	結果とその検証	48
7. 3. 1	推薦精度の検証	48
7. 3. 2	入力コストの検証	49
第8章 結論と今後の課題		53
8. 1	推薦精度について	54
8. 2	入力コストについて	54
8. 3	今後の課題	54
8. 3. 1	使用した履歴データについて	54
8. 3. 2	嗜好空間形成アルゴリズムについて	55
8. 3. 3	推薦アルゴリズムについて	56
参考文献		57

第 1 章

序論

1. 1 背景

近年、情報処理技術の急速な普及・発展に伴い、我々はインターネットなどを通じて、大量の電子化された情報へのアクセスが可能になった。様々な分野で、インターネットが持つ迅速性、広域性、利便性などの利点が活かされたシステムが開発されるようになり、たとえば取引の分野では、回線を通じて自由に売買できる電子商取引が拡大するようになった。それに伴い、音楽や映画などの様々なコンテンツに関して、それぞれのコンテンツが持つ基本的な情報・属性だけでなく、周囲の評判・意見・感想などが提供・公開されるようになり、我々は、自分が、「気になったときに」、「いつでも」それらの情報を得ることが可能になり、また、得ることが可能な情報量も急増した。提供された情報から、自分の必要な情報のみを抽出することを考えたときに、一般論として、利用者は可能な限り短時間で片付けてしまいたいと考える。しかし、情報量の膨大さ・情報の更新が収束しないことなどが原因で、利用者が自分の嗜好にあうコンテンツを獲得することが困難になっている。

このような作業を支援するアプローチのひとつとして、「推薦システム」(Recommender System)が開発されている。その要素技術である協調フィルタリング(Collaborative Filtering)は研究、実用ともに幅広い範囲で活用されている。協調フィルタリングは、人間同士が日常的に行う「口コミ」の原理を利用し、以下のような手順で推薦システムを実装したものである。

1. システムを利用する利用者(活動利用者)は、自身の嗜好を評価し、システムに渡す。
2. システムは活動利用者の評価を元に、既存利用者の中から活動利用者と類似した嗜好を持つ利用者(類似ユーザー)を見つけ出す。
3. システムは類似ユーザーが好むアイテムを活動利用者に推薦する。

協調フィルタリングは、人間の心理的側面を対象にしたアプローチであり、それが故に、多種多様なアルゴリズムが提案されていて、いまだ決定的な手法は確立されていない。(協調フィルタリングの各種アルゴリズム、問題点に関しては、第2章で詳しく述べる。)

1. 2 本論文の構成

既に、第1章では、本研究の背景について述べた。

第2章では、協調フィルタリング技術の流れを述べる。

第3章では、本研究の着眼点と既存システムの問題点、本研究で用いる解析手法、本研究の目的について述べる。

第4章では、提案手法にいたった経緯について述べる。

第5章では、本研究の提案手法について述べる。

第6章では、本実験で用いるパラメータ、アルゴリズムを決定するための予備実験について述べる。

第7章では、提案手法の評価実験とその検討、考察について述べる。

最後に、第8章では、本研究の結論と今後の課題について述べる。

第2章

協調フィルタリング

協調フィルタリングでは、表 2-1 のような、各利用者がそれぞれのアイテムについて評価したデータセットを用いる。利用者は自身が「知っている」アイテムについてのみ評価をするため、全ての $P_{i,j}$ (評価点) に数値が格納されているわけではない。(数値が空の要素を欠損値と称する。)

表 2-1. データセット (P_{ij} : 評価点)

	アイテム x_1	アイテム x_2	...	アイテム x_m
ユーザー1	P_{11}	P_{12}	...	P_{1m}
ユーザー2	P_{21}	P_{22}	...	P_{2m}
.	.	.		.
.	.	.		.
ユーザーn	P_{n1}	P_{n2}	...	P_{nm}

第 1 章の中でも簡単に触れたが、協調フィルタリングは大きく、以下のような 3 つのプロセスに分けることができる。

1. データ収集 : 活動利用者の評価を収集する。
2. 類似ユーザー探索 : 活動利用者に嗜好が類似している利用者を探索する。
3. 推薦アイテム推定 : 類似ユーザーの評価から、推薦するアイテムを推定する。

以下、順を追って、それぞれのプロセスについて説明する。なお、本章の説明は論文[1]に従う。

2. 1 データ収集

まず、活動利用者は自身が知っているアイテム x_i について評価点 $P_{i,j}$ を入力する。

評価の仕方としては、

- 自身の嗜好の度合いを数値表現する。

SD法 (図 2-1) : 形容詞対を両端とするものさし上の、適当な数値を選択する方法

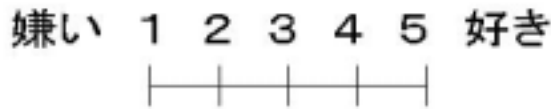


図 2-1. SD法 (Semantic Differential Technique)

- ・ 利用者が過去に取った行動（購買履歴など）を評価とする。

などの手法がある。

2. 2 類似ユーザー探索

続いて、活動利用者の評価をもとに、類似ユーザーを探索する。「類似ユーザーを探索する」と述べたが、アルゴリズムの視点から説明すると、厳密には「活動利用者と既存利用者との類似度 $sim(a,i)$ を算出する」という呼び方が適切である。

現実的に用いられるデータの場合、ほとんどの P_{ij} は欠損値であるため、類似度を算出する際には、活動利用者 a と既存利用者 i との共通要素のみが用いられる。例として、データが表 2-2 のような場合、利用者 1 との類似度はアイテム 1 とアイテム 2 に対するそれぞれの評価が、利用者 2 との類似度はアイテム 1 とアイテム 4 に対するそれぞれの評価が反映された類似度が算出される。利用者 3 との類似度は共通要素がないため、算出されない。

表 2-2. データセットの例

	アイテム1	アイテム2	アイテム3	アイテム4
活動利用者	$P_{a,1}$	$P_{a,2}$		$P_{a,4}$
利用者1	$P_{1,1}$	$P_{1,2}$		
利用者2	$P_{2,1}$		$P_{2,3}$	$P_{2,4}$
利用者3			$P_{3,3}$	

類似度という概念は、情報検索など多様な分野でも活かされ、様々な尺度が提案されている[2]が、協調フィルタリングで用いられる類似度は主に以下の2つである。

- ・ 相関係数

$$sim(a,i) = \frac{\sum_j (v_{a,j} - \bar{v}_a)(v_{i,j} - \bar{v}_i)}{\sqrt{\sum_j (v_{a,j} - \bar{v}_a)^2 \sum_j (v_{i,j} - \bar{v}_i)^2}}$$

- ・ コサイン尺度

$$sim(a,i) = \sum_j \frac{v_{a,j}}{\sqrt{\sum_{k \in I_a} v_{a,k}^2}} \frac{v_{i,j}}{\sqrt{\sum_{k \in I_i} v_{i,k}^2}}$$

$sim(a,i)$: 活動ユーザー a と既存ユーザー i との類似度 $v_{i,j}$: ユーザー i のアイテム j に対する評価点 \bar{v}_i : ユーザー i の評価点の平均 I_i : ユーザー i が評価したアイテムの集合
--

これらの尺度を元に、類似ユーザーを探索するわけだが、何千人、何万人という利用者が登録されたシステムの場合、類似度の高い上位 K 人のみを類似ユーザーとする方法がよく行われる。(K-nearest Neighbor Algorithm)

2. 3 推薦アイテム推定

最後に、活動利用者にとって「未知」のアイテムについて、潜在的嗜好予測値 $P_{a,j}$ を算出する。定式化すると以下のようなになる。

$$P_{a,j} = \bar{v}_a + \omega \sum_{i=1}^n sim(a,i)(v_{i,j} - \bar{v}_i)$$

システムは活動利用者に対して、数値 $P_{a,j}$ の高いアイテム j から順に推薦することになる。

第3章

既存システムの問題点と

本研究の目的

3. 1 本研究の着眼点

アイテムを評価する手順は推薦システムにとって非常に重要な要素である。なぜなら、利用者にとって面倒な処理になるほど、システムを用いることを放棄してしまう利用者を引き起こすことにつながり、一方で、活動利用者について十分な嗜好情報を集めることができなければ、それは結果的に推薦精度の低下につながるからである。そこで、本章では、2. 1節で述べた「データ収集」のプロセスについて考えていく。

3. 2 既存システムの現状分析と問題点

協調フィルタリングにおいては、様々な角度から問題点が指摘されているが、その中から、「データ収集」に基づく問題点を取り上げる。

3. 2. 1 新規利用者問題 (start-up 処理)

新規にシステムに参加する利用者の嗜好情報を収集する方針として、主に以下の2つが用いられている。

方針1：利用者に自由にアイテムを選んでもらい評価する。

方針2：システム側がアイテムを提示し、その中から利用者自身が「既知」のアイテムを評価する。

方針1を採っている映画推薦システムに **Amazon.com**[3]がある。このようなシステムでは、「映画タイトル」「監督」「出演者」などを入力する文字列検索や、映画の「ジャンル」「公開年」などを指定する絞り込み検索をさせ、活動利用者に自由に評価をしてもらう方針を採っている。このようなシステムでは、自身の印象に残ったアイテムを優先的に評価できるメリットがある反面、以下のようなデメリットがある。

Amazon.com の評価点の分布を表したものが図3-1である。図をみると大多数の利用者が、自身の好むアイテムばかりを評価しているということが一目瞭然であり、方針1では利用者の嗜好情報を十分に収集することができない。大多数の利用者が「好き」あるいは「嫌い」と評価するようなアイテムを活動利用者に評価させることは避けるべきである[4]。

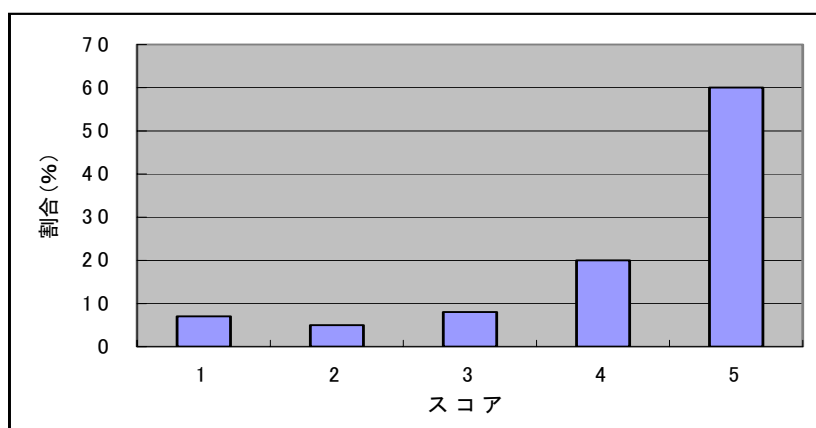


図3-1. Amazon.comによる利用者の評価分布

方針2を採っている推薦システムの代表例が、映画推薦システム MovieLens[5]である。まず、方針2の概要について MovieLens を例に簡潔に述べる。新たに参加した利用者に対して、システムはまず、図3-2のような画面を出力する。



図3-2. MovieLens 出力画面

活動利用者はシステムが提示する映画のリストから、見たことのある映画を探して、評価値を入力する。この処理は、活動利用者が評価値を入力した映画の合計が15に達するまで続けられる。この方針では、システムが提示するアイテムについて評価を行うため、利用者の本来の負担である「好きか嫌いか」の判定に加えて、「既知か未知か」の判定を行う必要がある。故に、利用者の処理コストが高くなり、システムの利用をあきらめてしまう利用者を引き起こす恐れがある。

一方で、システムの評価分布（図3-2）を見てみると、観測値の偏りが緩和されているのがわかる。結果的に、方針1に比べて活動利用者の嗜好情報をより詳細に収集することが可能であるというメリットがある。

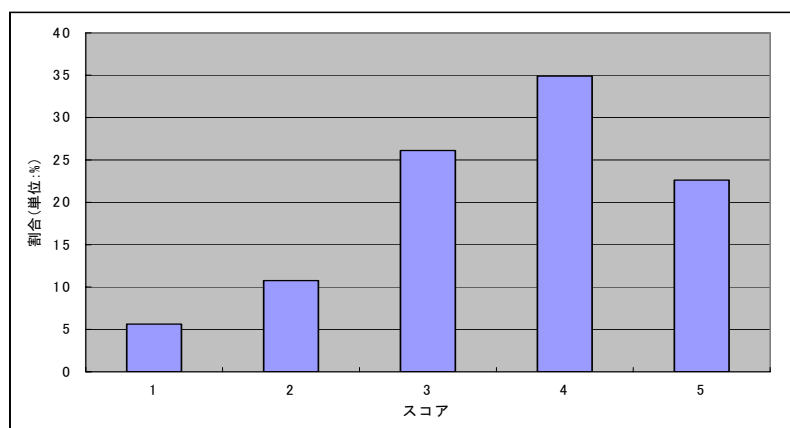


図3-2. MovieLens による利用者の評価分布

3. 2. 2 利用者の評価手法（SD法、順序応答）

第2章でも述べたように、ほとんどの推薦システムでは、利用者の評価手法にSD法を用いている。本来、SD法による計測は順序尺度であり間隔尺度ではなく、協調フィルタリングには適さないということが論文[6]で指摘されている。その概要を以下に述べる。

例えば、ある利用者が2つのアイテムに対して、それぞれ5段階中、「1」、「5」と評価したとき、利用者は「とても好き」、「とても嫌い」という意思表示をただけであって（順序尺度）、「一方はもう一方より5倍好き」という評価を下したわけではない（間隔尺度）。しかし、協調フィルタリングにおいては解析手法の不足から、類似度算出や潜在的嗜好予測値算出において、観測値を「間隔尺度」として用いている。そのため、

- ・ 被験者ごとに目盛の間隔は等しい
- ・ 全被験者の間で尺度の全長は一致する

という非現実的な仮定が導入される[7]。この仮定の下では、観測値は「中間」を中心とする対称な分布となっていることが期待されるが、現実には図3-1や図3-2のように非常に偏った分布となっているため、利用者の正確な計測は阻害されている。

以上のような理由から、論文[6]ではSD法の代替案として、図3-3のような「順序応答」による評価手法を提案し、協調フィルタリングの観測値として「順位」を適用した。その結果、寿司の嗜好データを対象として、推薦精度がSD法に基づくものよりも高いことを示した。



図3-3. 順序応答

第一の問題点として、順序応答は映画推薦システムのような大規模なデータセットへの適用が困難であることが挙げられる。一人の利用者が既知のアイテムについて整列することが可能なのはせいぜい10程度である。大量のアイテムが存在し、また、アイテムが日々更新されていくデータセットへの適用は難しいといわざるを得ない。

第二の問題点として、利用者は間隔尺度を強調できないという点が挙げられる。たとえば、利用者が図3-3のような評価を下した場合、ぶり⇔まぐろ、まぐろ⇔あじの嗜好度合いの差は同じものとして扱われる。「ぶり以外は嫌い」という評価も「4つ全てが好きだ」という評価も順序応答の下ではまったく同じものとして扱われるため、利用者の嗜好の計測という点で、依然として不備があることは否めない。

3.3 本研究の目的

本研究では以下の2つの目的を掲げる。

3.3.1 活動利用者の入力コストの軽減

「データ収集」プロセスで重視すべき要素は、「利用者の負担を軽減すること」「できるだけ利用者の嗜好を特徴付けること」の2つである。両者は本章のはじめに述べたようにトレード・オフの関係にある。とはいえ、推薦システムを支えているのはあくまでもシ

システムを訪れた既存利用者の「履歴」であり、利用者にシステムの利用を放棄させることはシステムを洗練させていく上で非常に「損」であると筆者は考えている。

そこで、本研究では「活動利用者の入力コストの軽減」を重視し、**3. 2節**で述べた問題点を踏まえた上で、「SD法」「順序応答」に替わる嗜好表現手段を提案する。（ここでいう「入力コスト」とは、処理に費やした時間などの実測的なものだけでなく、「疲労度」や「モチベーションを保てたかどうか」など心理的な側面も含む。）さらに、その嗜好表現手段にあわせた推薦アルゴリズムを実装し、従来の協調フィルタリングアルゴリズムに比べて推薦精度を落とさないことを示すことを目的とする。

3. 3. 2 「見たい」映画を推薦するシステム

映画や音楽などのコンテンツを利用者が欲しがる場合、本来、その人のその日の気分が大きく関わってくるものである。ところが、協調フィルタリングを主として「活動利用者の嗜好」に合わせたアイテムを推薦するシステムは数多くあるが、「活動利用者が現在見たい映画」を推薦するシステムは現状では皆無である。そこで、「活動利用者が現在見たい映画」を推薦することを可能にするアルゴリズムを構築することも合わせて本研究の目的とする。

第4章

本研究の経緯

4. 1 本研究で用いる学習データ

本研究の経緯について述べる前に、本研究で用いる実験データについて述べる。本研究では、GroupLens[8]より公開されている、映画推薦システムMovieLens[5]の嗜好履歴データを用いる。主な特徴は以下のとおりである。

- ・ 既存利用者数は6040人。利用者に付与されている属性は「性別」、「職業」、「年齢層」の3つであり、全ての利用者が、20タイトル以上の映画について評価をしている。
- ・ 登録されている映画数は3952タイトル。付与されている属性は公開された「年」または「年代」、「ジャンル」の二つである。

本研究では第7章の「本実験」にいたるまで、一貫してこのデータを学習データとして用いた。

4. 2 本研究当初のアプローチ

まず、本研究当初のアプローチについて、基本的な手順の流れを以下に示す。図示したものを図4-1に示す。

1. 学習データを、次節で述べる数量化理論によって解析し、映画ごとの位置関係、および、利用者ごとの位置関係を二次元平面上に集約することで、活動利用者が各個人の位置関係を「視覚」によって捉えることを可能にする。
2. 活動利用者は、二次元平面マップ上に「自身の嗜好」の範囲、あるいは「利用者が現在見たい映画」の範囲を指定する。
3. システムは「活動利用者の嗜好範囲」と「既存利用者の嗜好範囲」の情報を用い、類似ユーザーを断定する。
4. 結果を、2.3節以下の従来の協調フィルタリングアルゴリズムに適用し、推薦する映画を特定する。

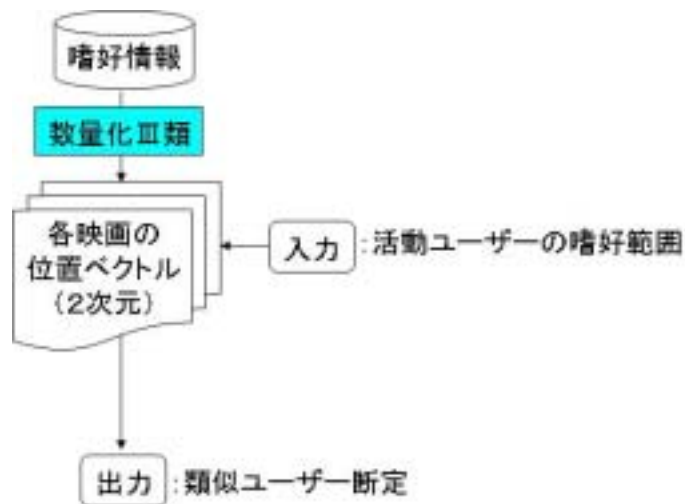


図4-1. 当初のアプローチ

協調フィルタリング技術においては類似ユーザーの特定のために、活動利用者はアイテムを一つ一つ評価しなければならない。この手法で利用者がアイテムを評価する際、**3.2節**で述べたような問題点がある。そこで、「活動利用者が類似ユーザーを直接指定することはできないだろうか、という考えがこのアプローチのもっとも重要な要素である。もちろん、活動利用者に類似ユーザーを選択させることは不可能なので、「お互いの嗜好範囲がどの程度の割合で交錯しているか」を類似度として用いることはできないか、と考えた。このアプローチが機能すれば、利用者の入力コストが大幅に軽減されることは明らかである。

ここで、視覚化するための解析手法として用いた数量化理論について概説する。

4. 3 数量化理論

4. 3. 1 数量化Ⅲ類

表 4-1. 利用者-アイテム関係行列

利用者	アイテム					合計
	1	...	j	...	m	
1	d_{11}	...	d_{1j}	...	d_{1m}	f_1
⋮	⋮	...	⋮	...	⋮	⋮
i	d_{i1}	...	d_{ij}	...	d_{im}	f_i
⋮	⋮	...	⋮	...	⋮	⋮
n	d_{n1}	...	d_{nj}	...	d_{nm}	f_n
合計	g_1	...	g_j	...	g_m	$2\sum_{i=1}^n \sum_{j=1}^m d_{ij}$

いま、表 4-1 のデータ行列のように、 n 人の利用者と m 人のアイテムとの関係が与えられているものとする。ただし d_{ij} は利用者 i がアイテム j を好む確率とする。このとき、各利用者、各アイテムにデータ行列に基づく相関係数が最大となるようにパラメータを与え、より趣味の類似している利用者同士がより近い座標に、より傾向の似ているアイテム同士がより近い座標となるようにパラメータを与えるのが数量化Ⅲ類の目的である。

利用者に与えるパラメータを $\vec{a} = [a_1, a_2, \dots, a_n]^T$ 、

アイテムに与えるパラメータを $\vec{b} = [b_1, b_2, \dots, b_m]^T$

とすると、 \vec{a} 、 \vec{b} は、

$$\begin{aligned} \max_{\vec{a}, \vec{b}} \quad & \vec{a}^T U \vec{b} = 1 \\ \text{subject to} \quad & \vec{a}^T U_x \vec{a} = 1 \\ & \vec{b}^T U_y \vec{b} = 1 \end{aligned}$$

$$\text{ただし、} \begin{pmatrix} U = [d_y] \\ U_x = \text{diag}(f_i) \\ U_y = \text{diag}(g_j) \end{pmatrix}$$

の解で与えられる[9]。この条件付き最適化問題は、Lagrangeの未定乗数法により、上式のLagrange関数は、

$$L = \vec{a}^T U \vec{b} - \frac{\lambda_1}{2} (\vec{a}^T U_x \vec{a} - 1) - \frac{\lambda_2}{2} (\vec{b}^T U_y \vec{b} - 1)$$

となる。この関数を整理すると、 $n \times n$ の対称行列 $U_x^{-\frac{1}{2}} U U_y^{-1} U^T U_x^{-\frac{1}{2}}$ の固有値問題に帰着させることができる（ただし、 $n \leq m$ とする）。従って、最適パラメータ \vec{a} 、 \vec{b} は行列 $U_x^{-\frac{1}{2}} U U_y^{-1} U^T U_x^{-\frac{1}{2}}$ の固有ベクトルにより与えられる。しかし、確率行列の固有値に関する一般論から、行列 $U_x^{-\frac{1}{2}} U U_y^{-1} U^T U_x^{-\frac{1}{2}}$ の最大固有値は常に1となり、固有値1のときのパラメータは $a_1 = \dots = a_n = b_1 = \dots = b_m$ となるため[9]、2番目に大きな固有値を第1軸として用いることになる。

数量化Ⅲ類、および次に概説する数量化Ⅳ類は、主にデータ行列で表現された近接関係を保ちつつデータを低次元空間に配置する多変量解析法として用いられ、特にデータ構造を要約し、視覚化する解析手法として用いられる。二次元にする場合、第2固有値 λ_2 と第3固有値 λ_3 に対応する固有ベクトルをそれぞれ分散1に標準化し、利用者とアイテムの二次元座標値を得ることができる。

また、このような低次元に情報を要約する解析手法においては、元のデータの情報量のうち、どれくらいの割合を各次元で表現することができたのかを判断するために「寄与率」と呼ばれる尺度を用いる。第 i 軸の寄与率は、次の式で表現される。

$$\frac{\lambda_i}{\sum_{i=1}^n \lambda_i}$$

さらに、複数低次元に要約するときには「累積寄与率」が用いられ、二次元の場合、次の式で表現される。

$$\frac{\lambda_1 + \lambda_2}{\sum_{i=1}^n \lambda_i}$$

4. 3. 2 数量化IV類

数量化IV類では、データ間の類似性に基づいて、似ているもの同士が近い値、似ていないもの同士が離れた値になるように、各個体に座標を与えることを目的とする。データ i と j の類似度を $sim(i, j)$ 、データ数 n 、与えるパラメータを $\vec{a} = [a_1, a_2, \dots, a_n]^T$ とすると、以下の最適化問題に定式化でき、

$$\begin{aligned} \max_{\vec{a}} \quad & - \sum_{i \neq j} sim(i, j) \times (a_i - a_j)^2 \\ \text{subject to} \quad & \sum_{i=1}^n a_i = 0 \\ & \sum_{i=1}^n a_i^2 = 1 \end{aligned}$$

を得る。

この最適化問題を解くことにより、 $n \times n$ の対称行列 $H = [h_{ij}]$ の固有値問題に帰着化できる。

$$h_{ij} = \begin{cases} \text{sim}(i, j) & (i \neq j \text{ のとき}) \\ - \sum_{j=1, \neq i}^n \text{sim}(i, j) & (i = j \text{ のとき}) \end{cases}$$

4. 3. 3 特異値分解

$n \times m$ 行列 U を、 q 次元の行列 $U_{(q)}$ に近似する手法として、特異値分解がある。

行列 U は

$$U_{n \times m} = A_{n \times r} D_{r \times r} B_{r \times m}^T$$

のように分解できることが知られている。 A, B は正規直交ベクトルを列ベクトルに持つ行列 ($A^T A = B^T B = I$)、 D は $\lambda_1, \dots, \lambda_r$ を対角要素に持つ対角行列である。上式は行列 U の特異値分解、 $\lambda_1, \dots, \lambda_r$ は特異値と呼ばれる。

上式より、

$$\begin{aligned} U^T U &= (A^T D B)(A D B^T) = B D^2 B^T \\ U U^T &= (A D B^T)(A^T D B) = A D^2 A^T \end{aligned}$$

となるが、このとき、 $D^2 = \{\lambda_1^2 \dots \lambda_r^2\}$ は $U^T U$ と $U U^T$ の共通の固有値、 B の列ベクトルは $U^T U$ の固有ベクトル、 A の列ベクトルは $U U^T$ の共通の固有ベクトルとなり、固有値問題に帰着することができる。

このとき、 $(q+1)$ 番目以下の特異値 $\lambda_{q+1}, \dots, \lambda_r$ が小さければ、行列 U を

$$U_{(q)} = A_{(q)} D_{(q)} B_{(q)}^T$$

により近似したとしても、その誤差は小さいと考えられている。

以上のように、特異値分解は大量データをもつ行列を次元圧縮させるための解析手法として用いられている。

4. 4 映画情報を視覚化した前例

表2-1のような利用者-アイテム行列を視覚化した前例として、スプリングモデルなどの配置法を用いた手法[10]、ファジークラスタリングを用いた手法[9]がある。[9]では、利用者と映画を点として、評価値が与えられている利用者と映画とを辺で結び、辺の重みを評価値としてクラスタリングを行っている。その結果、活動利用者の位置から、最も近い位置にある映画を推薦するというものである。

この手法の第一の問題点として、まず、SD法による利用者の評価値を利用者の映画に対する重みとして扱っている点が挙げられる。例えば、5段階評価をシステムが求め、利用者が「3」を評価値としたとすると、利用者は「どちらでもない」という評価をただけであって、「その映画を好む確率が50%である」という評価をしたわけではない。例えば、「3」をつけた映画も「1」をつけた映画も利用者側から見れば、ほとんど「無価値」なのかもしれない。また、3.2.2節で述べた主張からも、SD法による評価値を間隔尺度として用いることは得策ではなく、この手法から得られる利用者間、および映画間の類似性は妥当とはいえない可能性がある。

第二の問題点として、クラスタリングは本来、類似性の高いアイテム同士をグループ化する手法であるという点が挙げられる。映画同士の類似性は、映画の「人気」属性による側面が大きく、人気のある映画は一定範囲内に集中する傾向が強い（後に、本研究が数理化理論を用いてクラスタリングした結果を示す）。クラスタリングの位置情報を用いて、自身の位置から近い映画を推薦すると、ほとんどの利用者に対して、人気のある映画ばかりが推薦されることになる。賢明な推薦システムであれば、活動利用者の嗜好に合う、人気のない映画も推薦されるべきである。（ただし、[9]はあくまで、クラスタリングのみを目的とした研究である。）

4. 5 アプローチの検証

本研究では、類似する個体のまとまり具合を推定する解析手法（クラスタリング）として、数量化Ⅲ類と特異値分解を用いた。そこで、検証に入る前に、このような解析手法が数多く存在する中で、特にこの二つを用いた理由について述べる。一般的に、クラスタリングには数量化Ⅲ類よりも、数量化Ⅳ類の方が性能がよいといわれている[9]。しかし、数量化Ⅳ類を適用しようとする、類似度を定義する必要がある。ところが、今回用いたデータはSD法による5段階評価で記述されているため、3.2.2節で述べたような理由から間隔尺度として用いることは得策ではなく、また5段階評価であるため当然順序尺度として用いることもできない。故に、元のデータから類似度を抽出することは困難であり、また、本研究でおよそ4000タイトルもの映画の類似度を定義するのに十分なデータを

収集することは不可能である。元のデータから抽出することが可能なのは、「その映画を好きか嫌いか」という評価だけである。以上のような理由から、本実験では、0-1データに対応できる数量化Ⅲ類および特異値分解を選択した。

ここで、元のデータの観測値を変換するかどうかについては意見の分かれるところである。SD法による評価の観測値は一般的に、「評価軸の端をよく用い、明確にしようとするタイプ」「真ん中付近をよく用い、どっちつかずの評価が多いタイプ」と評価値の分布に個人差がある。（実際、本研究で用いたデータにはひとつも好きな映画がないとしている利用者が2人存在する。）協調フィルタリングでは、第2章の与式を見てもらえれば判るように、各個人の平均値を基準値とすることで、個人差を解消しようとしている。しかし、中心（今回の場合は「3」）を基準として好き嫌いを判断してくださいという質問の仕方をしている以上、好きなものがひとつもない利用者は意識的に嫌いな映画ばかりを評価し、好きな映画を避けているケースも考えられる。そこで、好きな映画から嫌いな映画までまんべんなく評価をして下さい、という質問の仕方をしていない以上、ばらつきの大きさに関わらず、利用者の評価値を尊重すべきであると本研究では考えた。よって、元のデータの観測値が「4」「5」の映画を、そのまま直接、利用者が好きな映画であるとして、本研究では用いた。

数量化Ⅲ類は、本来、各個体の位置情報を視覚的に捉えるために用いられる解析手法であるが、今回のような大規模なデータセットを適用した場合、アイテムの特徴に応じたいくつものクラスターが形成されることが期待される。二次元平面状に要約された情報をもとに、図4-2のように、活動利用者が自身の推薦して欲しい映画情報の範囲を指定することができれば、「活動利用者の入力コスト」は大幅に軽減することが可能であるといえる。

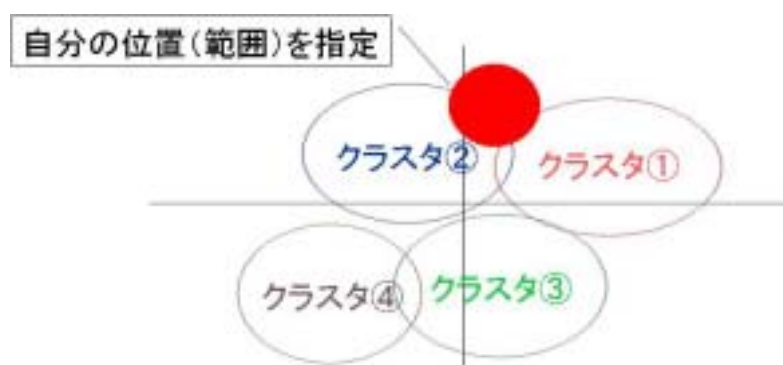


図4-2. 二次元平面上への範囲指定

ところが、結論から述べると、このアプローチから利用者が自身の嗜好の範囲を特定するのに十分なクラスター情報を獲得することはできなかった。その最大の要因は、元のデータが持つおよそ4000次元の情報を説明できる軸がとれず、二次元平面上に要約し切れなかったことにある。図4-3～図4-5に本研究で試みたクラスタリングの結果を示す。

図は、配置された各映画の位置座標に対して、その映画が持つ属性に基づいて色付けしたものである。図4-3は「人気」属性に基づく結果、図4-4は「公開された年」属性に基づく結果、図4-5は「ジャンル」属性に基づく結果である。

図4-3を見ると、人気のある映画ほど中央付近に集まる傾向があるといえる。図4-4を見ると、おおよそx軸を境として1995年以前の映画と、それ以後の映画が分離されている様子が見て取れ、古い映画ほどy座標の値が高そうなのがわかる。ところが、図4-5を見ると、ジャンルによってどういう傾向があるのか、まったく読み取ることができない。利用者が映画の好き嫌いを判断する場合、「ジャンル」属性によることが多く、この図を用いて、嗜好の範囲を特定するのはあまりにも無理がある。

また、結果の当該2固有値の累積寄与率は2%足らずであった。統計的尺度の面からもデータが二次元に集約し切れていないことは明らかである。

以上のような理由から、本研究では、

映画推薦システムで扱う大規模データが持つ情報を二次元平面上に集約することは困難である

と結論付け、本章で述べたアプローチを断念した。そこで、本研究では、本章の解析により得られた位置ベクトルを多次元用いることにした。

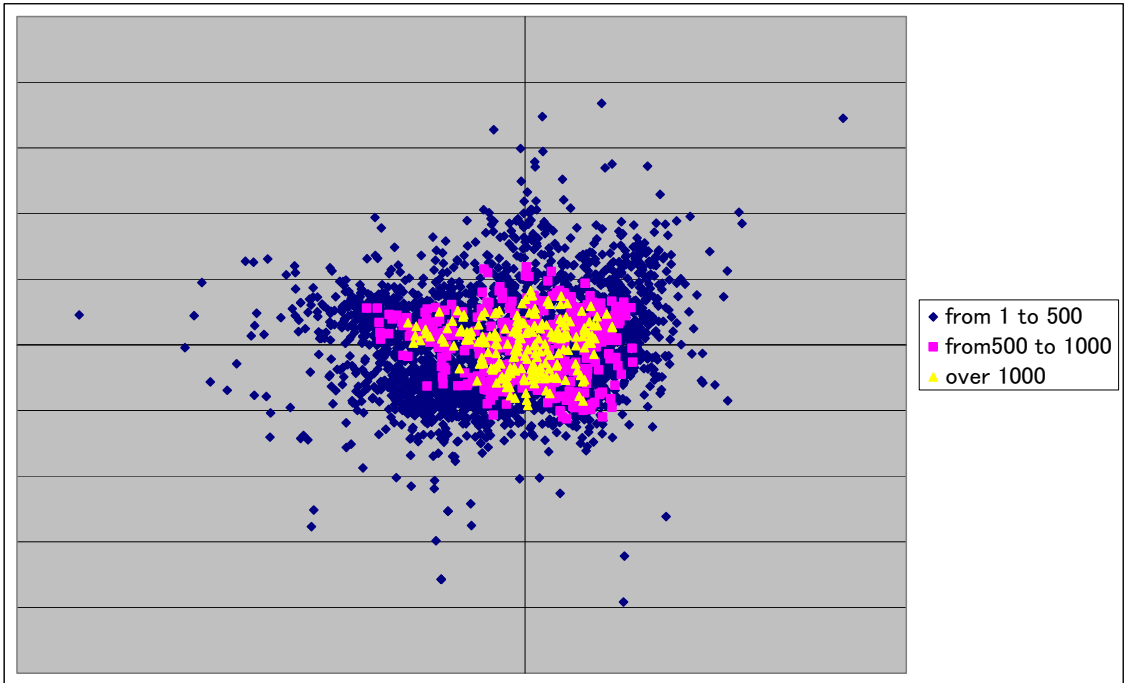


图 4 - 3 . 人气别

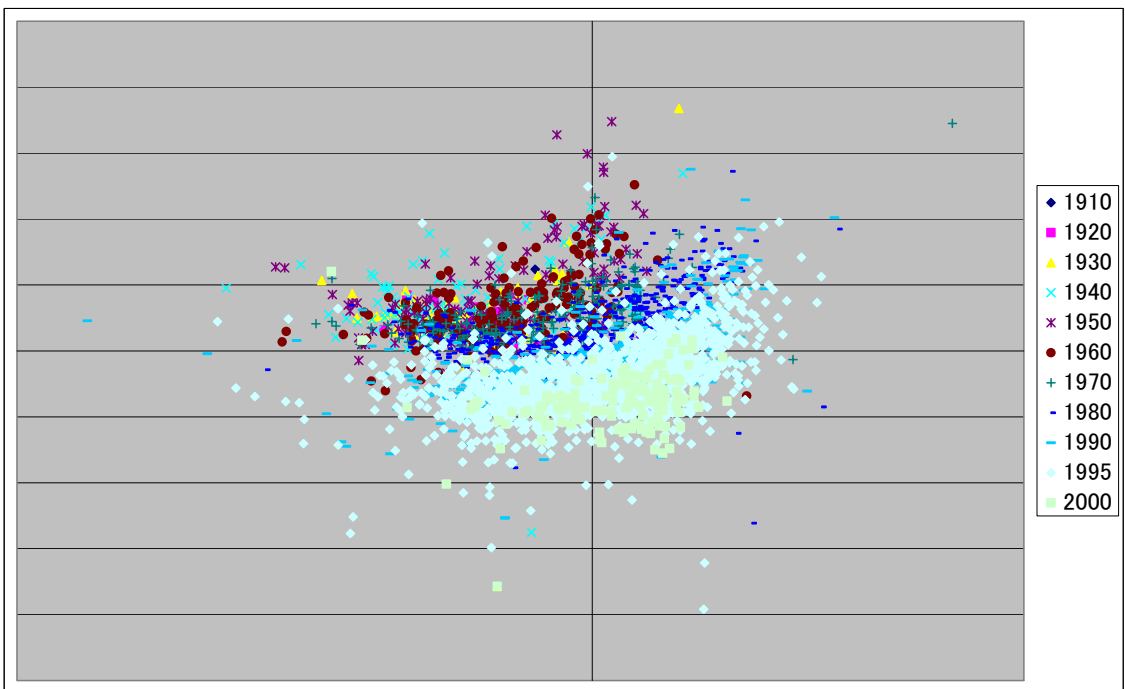


图 4 - 4 . 年度别

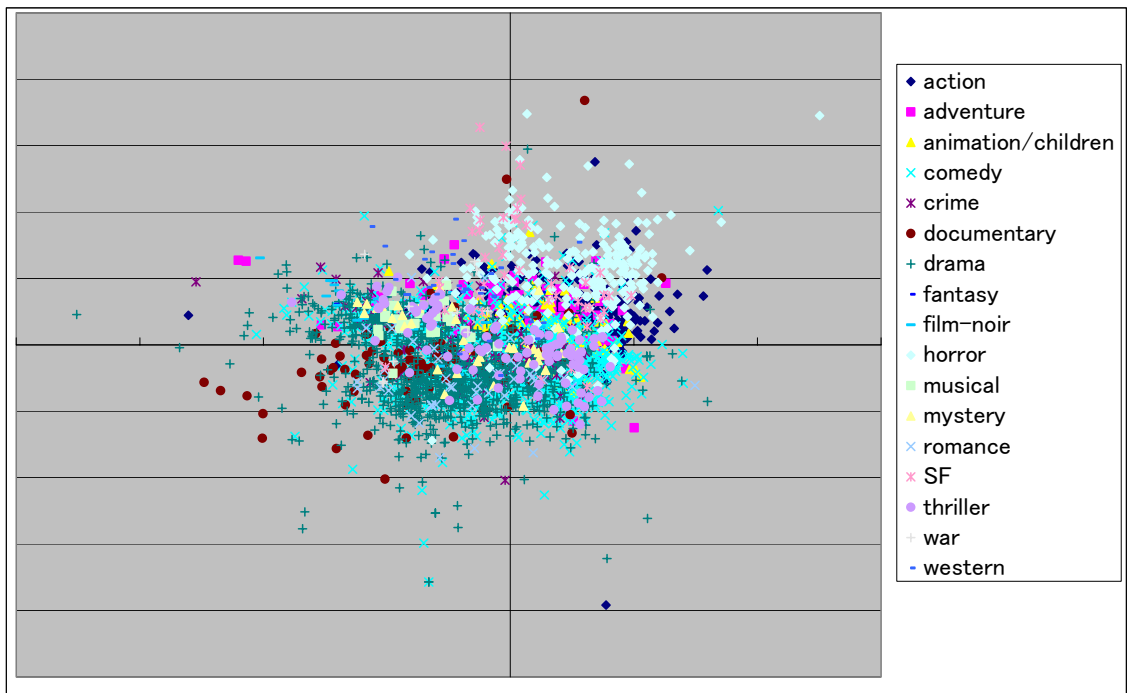


図4-5. ジャンル別

第 5 章

提案手法

本研究で提案する推薦アルゴリズムは、協調フィルタリングとほぼ同じ手順を取る。ただし、推薦する映画を決定する手法に類似ユーザーの情報は用いず、第4章のアプローチにより形成された映画の空間配置を用いる。詳細は以下に示す。まず、提案手法の流れを以下に示し、また図示したものを図5-1に示す。

提案手法の流れ

1. 元のデータに数量化Ⅲ類、又は特異値分解を適用し、各映画の位置ベクトルを求める。
2. 活動利用者は映画同士を対とする物差しから、「好ましい」と思われるほうの映画を判断し、その度合いに応じた評価値を選択する。
3. システムは活動利用者の評価値から、アイテム空間上の活動利用者の嗜好範囲を特定する。
4. システムは、活動利用者の嗜好範囲に最も近傍にある、活動利用者にとって未知の映画を推薦する。

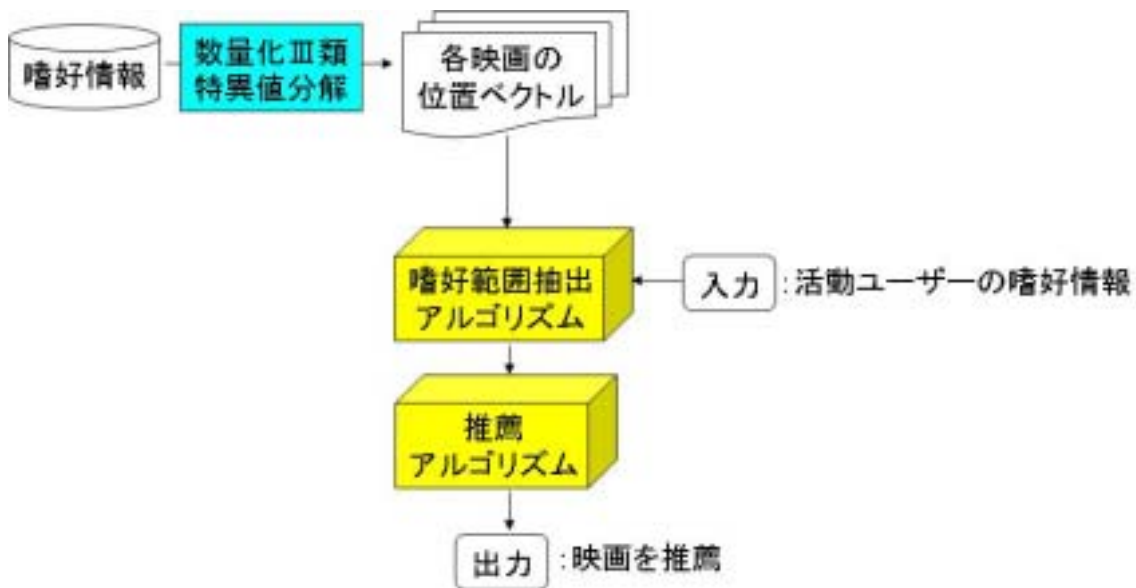


図5-1. 提案手法の流れ

以後、この手順に沿って提案手法について具体的に述べるが、その前に映画間の類似性を用いる手法の利点について述べる。類似ユーザーの情報を用いず、映画同士の類似性を用いた推薦を行うアプローチの前例として、[11]がある。利用者間の類似性は、活動利用者の評価が終わった後でなければ算出することができない。しかし、(利用者数を n とする

と) 計算量は $(n-1)$ で済む。映画間の類似性は活動利用者が評価を行う以前から算出が可能であるので、事前に用意できる。一方で、全ての組み合わせについて類似度を用意していなければならないため、計算量は二乗オーダー $(\frac{n(n-1)}{2})$ となる。故に、メモリを余計に必要とするが、リアルタイムの推薦を想定するならば、事前に類似度を用意できる後者のほうがよい。

手順1については既に第4章の中で述べた。そこで、手順2から本研究の考え方について述べる。

5. 1 データ収集

本研究では、利用者の嗜好表現手段として、SD法の代わりに、図5-2のような異なる映画同士を対とするものさしから評価値を選択する手法(一対比較法)を採択した。



図5-2. 一対比較法とSD法

一対比較法の最大の特長は、観測値を間隔尺度として用いることが可能であるという点にある。第2章で述べたように、SD法による観測値が、映画aが「4」、映画bが「1」であったときに、「その利用者は映画aを映画bと比較して4倍好んでいる」とする解釈は間違いである。しかし、一対比較法による評価値の距離比が1:4であった場合、「その利用者は映画aを映画bと比較して4倍好んでいる」とする解釈をすることは可能である。

ところが、観測値が一対比較法によるものであるとき、これを低次元のモデルで処理しようとする矛盾が起こりうる[12]。このことを文献[12]から引用し、ここで概説する。A, B, Cという3つのアイテムが存在するとして、ある利用者が $A > B$ (AはBより好ましい)、 $B > C$ という評価をしたとする。ところが、 $A > C$ という評価になるとは限ら

ず、 $C > A$ となることも十分に起こりうる。このことについて検討する。3つのアイテムが図5-3のように配置されているとする。

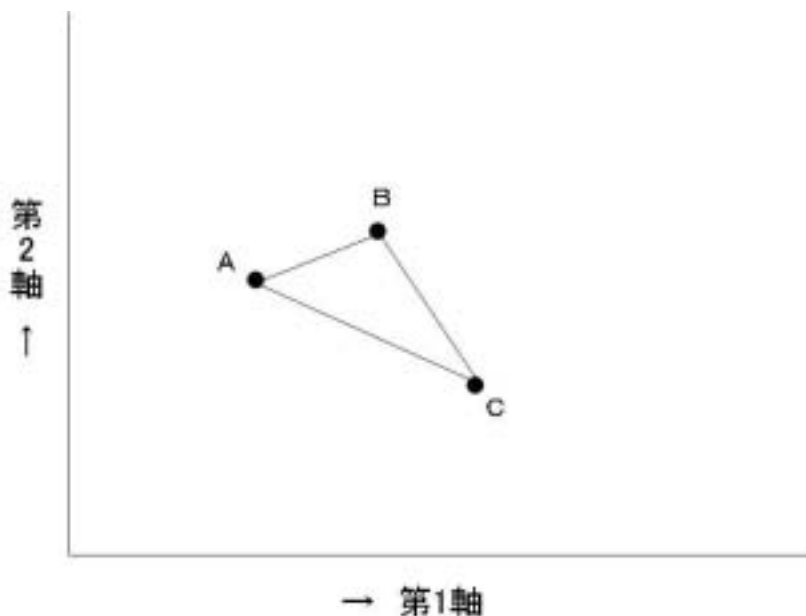


図5-3. アイテムの配置

図5-3は原点に近いものほど、利用者にとって好ましいものとして配置されているものと仮定する。このとき、 $A > B$ 、 $B > C$ という評価は第1軸の要因にしたがったものだと考えることができる。しかし、一見矛盾に思える、 $C > A$ という評価は第2軸の要因に従ったものであると考えることができる。仮に $B > A$ という評価が行われたのだとしたら、第1軸と第2軸だけでは説明がつかない。しかし、このような場合は、視覚化されていない第3の要因に従った判断であると考えれば妥当と考えられる。このように、一対比較法によって起こりうる矛盾は、各アイテムの高次元空間での配置と判断者の次元選択によって起こっていると考えられる。

5. 2 活動利用者の嗜好範囲の特定

次に、一対比較によって得られた観測値から、いかにして活動利用者の嗜好範囲を特定するかについて解説する。

いま、 m タイトルの映画について、それぞれの位置情報が n 次元ベクトルで表現されているものとする。位置ベクトルは元のデータを数量化Ⅲ類、または特異値分解で解析したものをを用いる。総数 n の各次元が映画の持つ各属性の要因であると考えれば、映画ごとの各次元のユークリッド距離は、要因ごとの映画間の非類似度であると考えて差し支えないであろう。なぜなら、6040人もの既存利用者の意見が反映された結果であるからで

ある。

さて、活動利用者は図5-4のように、映画Aと映画Bについて $p:(1-p)$ に分割する位置を一対比較により選択したものとする。

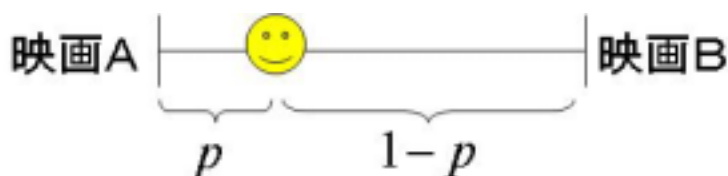


図5-4. 一対比較

このとき、映画Aと映画Bの位置ベクトルを $p:(1-p)$ に分割する「点の集合」を活動利用者の嗜好範囲とするのが提案手法のもっとも単純な考え方である。もし、 n 次元空間を可視化できていたなら、活動利用者は映画Aと映画Bの位置を $p:(1-p)$ に分割する位置を自身の嗜好の位置として指定したであろうとみなすのである。

一対比較法を用いた意思決定法として代表的なものにAHP（階層化意思決定法）がある[13]。本研究の推薦アルゴリズムの考え方は、一対比較の距離比をアイテムの重みとする点で、AHPと非常に似ている。最も異なる特徴は総評価数の違いである。AHPでは、まず、評価基準について何を重要視するのかを、評価基準同士の一対比較によって求める。映画で考えれば、「人気度」「ジャンル」「値段」などが評価基準にあたるだろう。そして、評価基準ごとのアイテム同士の一対比較を行う。評価基準数を n 、総アイテム数を m とすると、AHPの総評価数は

$$\frac{n(n-1)}{2} + n \times \frac{m(m-1)}{2}$$

となるため、大規模なデータセットに対して用いることは困難である。また、AHPはその解析手法の特徴から、利用者にとって未知のアイテムを織り交ぜることができない。AHPは、購買目的の分析などによく用いられる手法である。

提案手法では、アイテムの全ての組み合わせについて評価をするわけではない。また、評価基準ごとの一対比較はしない。しかし、評価基準のウェイトは用いる。数量化理論による解析から得られた位置情報の第1軸から第 n 軸をそれぞれ n 個の評価基準と仮想的にみなし、各軸の寄与率をウェイトとして用いる。

話を戻し、活動利用者の嗜好範囲特定のプロセスの説明に入る。簡単のために、 $n=2$ の

とき、映画Aと映画Bが図5-5、図5-6のように配置されているものとしよう。既に述べたように、各軸の位置は映画の属性に基づく位置を示していると考え。 (実際には点の集合は図に示すように垂直にはならないが、イメージとして示す。)

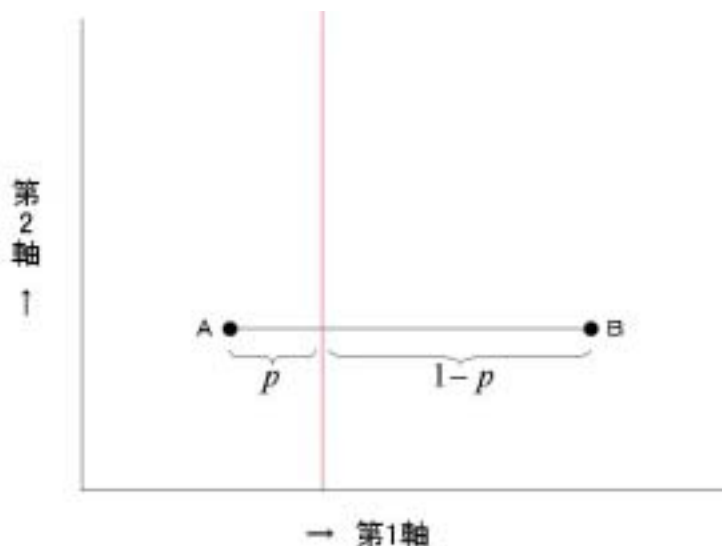


図5-5

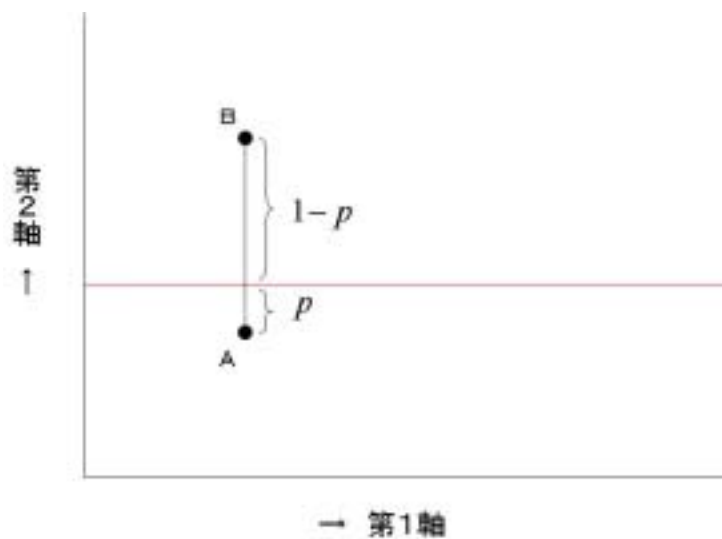


図5-6

提案システムは、活動利用者が選択した映画の対に応じて、映画の「属性」に対して評価をしていると考える。図5-5のように、映画対が第1軸の属性に著しく差のある組み合わせであれば、利用者は第1軸の属性に対して評価をしたとみなし、図5-6のように、第2軸に差のある組み合わせであれば、第2軸の属性に評価をしたと考える。結果として、

近似的には映画同士の各軸の距離差に応じた評価が行われたとみなすのである。

まず、利用者の嗜好の位置を、映画間を分割する「点」ではなく「点の集合」とした理由について述べる。本研究は「活動利用者の入力コスト削減」が目的のひとつであるため、嗜好情報の収集は、**3. 2. 1節**で述べた方針1をとることを想定している。このとき、たいていの場合、人気のある映画の対が選択されることになる。すると、**第4章**で述べたように、「点」の周りには人気のある映画ばかりが位置づけられている可能性が高い。故に、「点」から距離の近い映画を推薦することは、結果的に人気のある映画ばかりが推薦されることになることは容易に想像できるため、人気のない映画も網羅的に探索できるように、本研究では利用者の嗜好の位置を「点の集合」とした。

n 次元空間に話を戻す。活動利用者が k 通りの映画対について評価を行ったとする。このとき、 \overline{Mu}_i を活動利用者 i 番目の一対比較で使用した映画の位置ベクトル、 u_i を利用者が対として用いた映画、 p_i を一対比較による比率であるとする、活動利用者の嗜好空間は以下の条件を満たす \vec{x} となる。

$$\begin{aligned} \|\vec{x} - \overline{M}_{\{u11\}}\| : \|\vec{x} - \overline{M}_{\{u12\}}\| &= p_1 : (1 - p_1) \\ \|\vec{x} - \overline{M}_{\{u21\}}\| : \|\vec{x} - \overline{M}_{\{u22\}}\| &= p_2 : (1 - p_2) \\ &\vdots \\ \|\vec{x} - \overline{M}_{\{ui1\}}\| : \|\vec{x} - \overline{M}_{\{ui2\}}\| &= p_i : (1 - p_i) \\ &\vdots \\ \|\vec{x} - \overline{M}_{\{uk1\}}\| : \|\vec{x} - \overline{M}_{\{uk2\}}\| &= p_k : (1 - p_k) \end{aligned}$$

5. 3 推薦アイテム決定

推薦アルゴリズムは単純で、各映画の位置ベクトルから前節で定義した嗜好空間までの距離を、活動利用者と各映画との類似度とする。 d_j を映画 j から嗜好空間までの距離、 \overline{x}_j を嗜好空間上の映画 j からの最短位置を示すベクトル($1 \leq j \leq n$)とすると、提案手法は d_j が最小となる映画 j を見出す最適化問題として以下のように定式化し、

目的: $\arg \min \|\vec{x}_j - \vec{M}_j\|$

$$\text{制約条件: } \begin{cases} \|\vec{x}_j - \vec{M}_{(u11)}\| : \|\vec{x}_j - \vec{M}_{(u12)}\| = p_1 : (1 - p_1) \\ \|\vec{x}_j - \vec{M}_{(u21)}\| : \|\vec{x}_j - \vec{M}_{(u22)}\| = p_2 : (1 - p_2) \\ \vdots \\ \|\vec{x}_j - \vec{M}_{(uk1)}\| : \|\vec{x}_j - \vec{M}_{(uk2)}\| = p_k : (1 - p_k) \end{cases}$$

全ての d_j を算出し、降順に整列することで達成される。(図 5-7)

$n = 2$ のときを例に、もう少し判りやすくすると、位置情報が図 5-7 のように与えられているものとする。図の赤線が利用者の嗜好範囲、黄色の点 (a, b) は利用者が比較した映画の位置、赤色の点 (1, 2, 3) が利用者にとって未知の映画を示している。

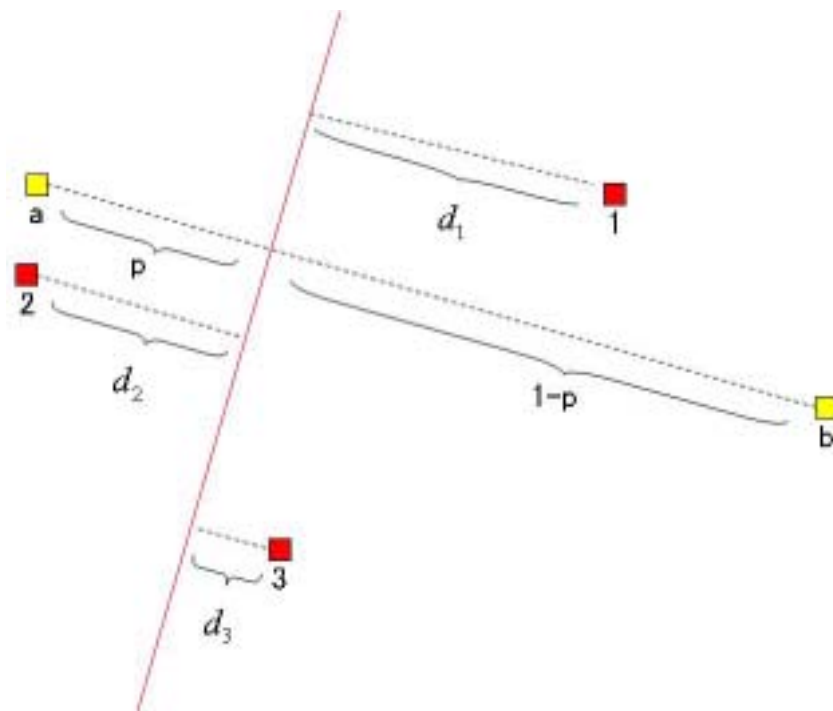


図 5-7. 提案手法のイメージ

この場合、 $d_3 < d_2 < d_1$ であるため、映画3 > 映画2 > 映画1の順で、利用者に対して推薦することになる。

ここで、制約条件の近似化について考える。上式をそのままLagrange関数に適用すれば、制約条件が二乗オーダーとなり、算出処理時間の面で活動利用者に負担がかかることは容易に想像できる。そこで本研究では、「映画間を $p:(1-p)$ に分割する点の集合」の代わりに「映画間を $p:(1-p)$ に分割する超平面」を採用することにした。超平面に近似化すれば、制約条件を一次元で表現することができる。 $n=2$ のとき、超平面は図5-5や図5-6で示したような垂直な直線そのものである。説明を判りやすくするために、あらためて、図5-8に示す。

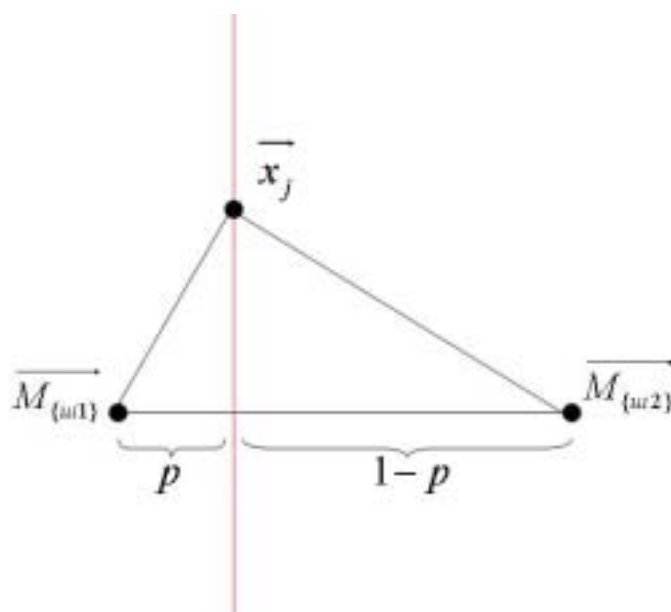


図5-8. 超平面

超平面上の点の集合 \vec{x}_j は内積の定理から、以下のような等式が成り立つ。

$$(\vec{x}_j - \vec{M}_{\{ui1\}}) \cdot \frac{(\vec{M}_{\{ui2\}} - \vec{M}_{\{ui1\}})}{\|\vec{M}_{\{ui2\}} - \vec{M}_{\{ui1\}}\|} = p \|\vec{M}_{\{ui2\}} - \vec{M}_{\{ui1\}}\|$$

整理すると、

$$(\vec{x}_j - \vec{M}_{\{ui1\}}) \cdot (\vec{M}_{\{ui2\}} - \vec{M}_{\{ui1\}}) = p \left\| \vec{M}_{\{ui2\}} - \vec{M}_{\{ui1\}} \right\|^2$$

$$\vec{x}_j \cdot (\vec{M}_{\{ui2\}} - \vec{M}_{\{ui1\}}) = \vec{M}_{\{ui1\}} \cdot (\vec{M}_{\{ui2\}} - \vec{M}_{\{ui1\}}) + p \left\| \vec{M}_{\{ui2\}} - \vec{M}_{\{ui1\}} \right\|^2$$

となる。上式を制約条件とすると、Lagrange定数 $\vec{\lambda} = \{\lambda_1 \cdots \lambda_k\}$ を導入し、Lagrange未定乗数法により解くために、

$$F = \left\| \vec{x}_j - \vec{M}_j \right\|^2 + \lambda_1 (\vec{a}_1 \cdot \vec{x}_j + b_1) + \cdots + \lambda_k (\vec{a}_k \cdot \vec{x}_j + b_k)$$

$$\text{ただし、} \quad \vec{a} = (\vec{M}_{\{ui2\}} - \vec{M}_{\{ui1\}})$$

$$\vec{b} = -\vec{M}_{\{ui1\}} \cdot (\vec{M}_{\{ui2\}} - \vec{M}_{\{ui1\}}) - p \left\| (\vec{M}_{\{ui2\}} - \vec{M}_{\{ui1\}}) \right\|^2$$

とおき、 F を x_j, λ_i について偏微分して0に等しく置く。すると次のように $n+k$ 連立一次方程式に展開され、これを解くことで提案手法の目的は達成される。

第6章

予備実験

6. 1 予備実験の条件

提案手法の検証に入る前に、本実験で用いる実験条件と各種パラメータの設定を行う必要がある。具体的には、

- ・ 各個体の位置情報を数量化Ⅲ類によるものにするのか、特異値分解によるものにするのか
- ・ パラメータ n (数量化Ⅲ類であれば第 n 軸まで用いるのか、特異値分解では q をいくつに設定するのか) の設定
- ・ パラメータ k (活動利用者にいくつの映画対を評価してもらうか) の設定

の3点である。

これらを決定するために、所属研究室の学生数名に協力してもらい、予備実験を行った。

6. 2 予備実験の内容

提案手法の検証のために、本研究では、数量化理論による解析は統計解析用フリーソフト「R」で、提案手法は「Java」でそれぞれ実装した。まず、実験手順について以下にしめす。

1. 活動利用者は、学習データに用意された3952タイトルの中から、見たことのある映画をいくつか選択する。
2. 活動利用者は、映画間のユークリッド距離の大きいものから順に50対の対比較を行う。
3. パラメータ k の値が、「10」、「20」、「30」、「40」、「50」のときの5パターンそれぞれについて最適化問題を解き、推薦される上位20タイトルずつ計100タイトルの映画について、公開されている映画データベースのあらすじや映画ジャンルなどの情報を参考に、活動利用者は評価する。

この実験を、映画位置情報が「数量化Ⅲ類によるもの」であるときと、「特異値分解によるもの」であるときの2パターン行ってもらった。 n は1000に設定した。この理由は**6.2節**の中で述べる。以後、それぞれの手順について、詳しく説明する。

6. 2. 1 映画の選択

まず、一対比較のペアを決めるために、表6-1に示す映画のリスト（左から順に「タイトル」「ジャンル」「人気順位」「チェックボックス」）を活動利用者に提示し、見たことのある映画についてチェックをつけてもらった。なお、利用者にできるだけ負担をかけないようにという配慮から、映画は人気のあるものから降順に整列されている。

表6-1. 提示映画リスト

1	アメリカン・ビューティー	Comedy Drama	1	0
2	Star Wars: Episode IV - A New Hope (邦名なし)	Action Adventure Fantasy Sci-Fi	2	0
3	Star Wars: Episode V - The Empire Strikes Back (邦名なし)	Action Adventure Drama Sci-Fi War	3	0
4	インディ・ジョーンズ:レイダース 失われたアーク	Action Adventure	4	0
5	プライベート・ライアン	Action Drama War	5	0
6	年たちの沈黙	Drama Thriller	6	0
7	マトリックス	Action Sci-Fi Thriller	7	0
8	シックス・センス	Thriller	8	0
9	Star Wars: Episode VI - Return of the Jedi (邦名なし)	Action Adventure Romance Sci-Fi War	9	0
10	ファード	Crime Drama Thriller	10	0
11	シンドラーのリスト	Drama War	11	0
12	ショーシャンクの空に	Drama	12	0
13	ターミネーター2	Action Sci-Fi Thriller	13	0
14	ゴッドファーザー	Action Crime Drama	14	0
15	ブレイブハート	Action Drama War	15	0
16	プリンセス・ブライド・ストーリー	Action Adventure Comedy Romance	16	0
17	バック・トゥ・ザ・フューチャー	Comedy Sci-Fi	17	0
18	恋におちたシェイクスピア	Comedy Romance	18	0
19	L.A. Confidential (邦名なし)	Crime Film-Noir Mystery Thriller	19	0
20	パルプ・フィクション	Crime Drama	20	0
21	マルコヴィッチの穴	Comedy	21	0
22	ジュラシック・パーク	Action Adventure Sci-Fi	22	0
23	ターミネーター	Action Sci-Fi Thriller	23	0
24	フォレスト・ガンプ	Comedy Romance War	24	0
25	恋はデジャブ	Comedy Romance	25	0
26	トイ・ストーリー	Animation Children's Comedy	26	0
27	メン・イン・ブラック	Action Adventure Comedy Sci-Fi	27	0
28	ET the Extra-Terrestrial	Children's Drama Fantasy Sci-Fi	28	0
29	エイリアン	Action Horror Sci-Fi Thriller	29	0
30	逃亡者	Action Thriller	30	0
31	ユー・ジュアル・サスペクツ	Crime Thriller	31	0
32	ゴーストバスターズ	Comedy Horror	32	0

↓

計3952タイトル

6. 2. 2 一対比較

まず、一対比較の組み合わせとしてお互いの位置間のユークリッド距離の遠い組み合わせを優先的に用いた理由について述べる。本アプローチの原点は「一対比較をした映画で囲まれる空間（本節に限り、 M_i とする）に、活動利用者の嗜好空間（同じく U とする）が存在する」という考えに基づくものである。つまり、 $U \subset M_i$ でなければならない。ところが、一対比較の組み合わせによっては $U \not\subset M_i$ となる可能性は十分に考えられ、このよ

うなケースは推薦精度の低下を導くことになる。よって、本研究では、ユークリッド距離の離れた組み合わせを選択し、できるだけ M_i の範囲を広げることでこの問題を回避することにした。

実験内容に話を戻す。活動利用者には、図6-1に示すように、Javaアプリケーション上で、自身の間隔尺度に基づく評価を行ってもらった。



図6-1. 一対比較用アプリケーション

6.2.3 推薦映画の評価

利用者の一対比較評価を受けて、システムは提案手法により、好ましいと思われる結果を利用者に返す。本実験では k の値を決めるために、「10」、「20」、「30」、「40」、「50」のときの5パターンそれぞれについて、20タイトルずつ結果を提示し、評価してもらった。

6. 3、実験条件の決定

6. 3. 1 解析手法

解析手法の選択は、最終的な推薦精度によって決定することにした。そこで、通常、協調フィルタリングにおいて、推薦精度の評価指標として用いられるMAE（平均絶対誤差：Mean Absolute Error）について説明する。

実験条件が平等である（入力評価数が同じ）とき、システムの比較に用いるMAEは以下の式で示される。

$$S_a = \frac{1}{m_a} \sum_{j \in P_a} |P_{a,j} - v_{a,j}|$$

$\left(\begin{array}{l} m_a : \text{推薦対象としたアイテム数} \\ P_a : \text{ユーザーが事前評価したアイテム数} \\ P_{a,j} : \text{潜在的嗜好推定値} \end{array} \right)$

MAEは推定値と実際値とのずれを示すものであり、統計的尺度としての意味はかなり高い。しかし、当然のことながら、本研究のように推定値と実際値が異なる尺度を用いている場合は使用できない。結局、本実験の推薦精度の比較は、推薦アイテムに対する利用者のスコアを用いるしか術がない。

表6-2に推薦されたアイテムに対する、パラメータ k ごとの各被験者のスコアを掲載する。

表 6 - 2. 各被験者の平均スコア

ユーザー 一対比較数	A		B		C	
	Ⅲ類	特異値	Ⅲ類	特異値	Ⅲ類	特異値
10	2.80	3.05	1.07	2.90	3.44	3.55
20	2.40	2.58	0.67	2.90	3.17	3.28
30	1.92	2.39	0.94	1.95	3.33	3.39
40	1.77	3.21	1.61	1.55	2.82	3.47
50	2.33	2.71	2.64	1.90	2.38	3.45
計	2.27	2.80	1.34	2.24	3.08	3.43

表に示されるスコアから、個人ごとのスコアのばらつきは大きいですが、全ての被験者が特異値分解による解析のほうが数量化Ⅲ類による解析よりも推薦精度の面で優れている。また、被験者ごとに、2つの標本の平均についてWelchのt検定を行うと、被験者Aの標本は5%の有意水準で有意な差があり、被験者Bの標本は1%の有意水準で有意な差があった。被験者Cの標本に関しては、両側確率が0.053となり、ほぼ5%の有意水準で有意な差があるといえる結果となった。さらに、利用者の意にそぐわないアイテムが推薦されてしまう確率についても検証してみる。(表6-3)

表 6 - 3. 被験者が嫌う映画が推薦される確率

A		B		C	
Ⅲ類	特異値	Ⅲ類	特異値	Ⅲ類	特異値
0.54	0.43	0.78	0.54	0.34	0.13

以上より、推薦結果に対する被験者のスコアからみると、特異値分解による解析手法のほうが予測精度の優れていることを示せた。

6. 3. 2 次元数 n

本実験では、 $n = 1000$ を用いた。パラメータ n は累積寄与率の面から決定した。 n を大きくしていったときの累積寄与率の推移を表 6 - 4 に示す。

表 6 - 4. 累積寄与率

n	数量化Ⅲ類	特異値分析
100	20.5	49.1
200	30.3	59.6
300	37.9	67.4
400	44.4	—
500	50.1	78.4
1000	70.9	92.2
1500	84.0	—
2000	92.1	—

要約したデータが、元の多次元データの変動を説明できているという意味で、累積寄与率はおおよそ 7 ~ 8 割程度は必要である [14] という点から、 $n = 1000$ という値に設定した。特異値分解による位置情報は、500 次元目以降はほとんど役に立っていないと考えられるが、ここでは実験条件に公平を帰する為、同値を用いた。なお、累積寄与率の差異も、推薦精度の優劣に影響した判断基準となりうる。

6. 3. 3 一対比較数 k

本節では、 k をいくつに設定するかについて考える。表 6 - 2 から、評価数に応じた推薦精度の傾向を得ることは到底できない。また、一対比較評価を 10 ずつ追加していった結果にもかかわらず、表 6 - 5 (被験者 C に対する提示リスト) のように推薦される映画が重複することはほとんど見られなかった。この原因について追究する。なお、本節で示す提示結果は全て、特異値分解の解析によるものである。

まず、 k を大きくすることによって起こりうるデメリットの可能性について言及する。本アルゴリズムでは、類似性の低い映画の組み合わせについて優先的に評価を行うため、 k が大きくなっていくにつれて、被験者の判断が難しくなることが予想される。表 6 - 6 は action 映画を推薦して欲しいと考えている被験者が、action 映画に極端に評価を寄せ、action 映画を含まない対に関してはすべて中点を選択したときのシミュレーションの結果である。

表6-5. 被験者Cに対する提示リスト

映画タイトル	映画ジャンル
k=10のときの推薦結果	
Aladdin (アラジン 特定不能)	Animation Children's Comedy
バグズ・ライフ	Animation Children's Comedy
わんわん物語	Animation Children's Comedy
メリーポピンズ	Children's Comedy Musical
マトリックス	Action Sci-Fi Thriller
美女と野獣	Animation Children's Musical
シンデレラ	Animation Children's Musical
リトル・マーメイド	Animation Children's Comedy
夢のチョコレート工場	Adventure Children's Comedy
ロジャー・ラビット	Adventure Animation Film-Noir
シックス・センス	Thriller
ピーターパン	Animation Children's Fantasy
カラー・オブ・ハート	Comedy
バンビ	Animation Children's
マミー	Action Adventure Horror Thriller
ナイトメア・ビフォア・クリスマス	Children's Comedy Musical
チキンラン	Animation Children's Comedy
ジャングル・ブック	Animation Children's Comedy
ウォレスとグルミット ペンギンに気をつけろ	Animation Comedy
Batman (バットマンのどれか)	Action Adventure Crime Drama
k=20のときの推薦結果	
アメリカン・ビューティー	Comedy Drama
フル・モンティ	Comedy
恋愛小説家	Comedy Drama
L.A. Confidential (邦名なし)	Crime Film-Noir Mystery Thriller
グッド・ウィル・ハンティング	Drama
カラー・オブ・ハート	Comedy
ザ・エージェント	Drama Romance
プリティ・リーグ	Comedy Drama
Dave (候補多数)	Comedy Romance
ポイント・ブランク	Comedy Crime
ワグザドッグ~ウワサの真相~	Comedy Drama
リプリー	Drama Mystery Thriller
Wedding Singer The (邦名なし)	Comedy Romance
go	Crime
マトリックス	Action Sci-Fi Thriller
エリザベス	Drama
チキンラン	Animation Children's Comedy
ファーゴ	Crime Drama Thriller
ボーイズ・ドント・クライ	Drama
市民ケーン	Drama

表6-6. アクション好きの被験者に対する提示リスト

映画タイトル	映画ジャンル
k100のときの推薦結果	
ターミネーター2	Action Sci-Fi Thriller
マミー	Action Adventure Horror Thriller
トータル・リコール	Action Adventure Sci-Fi Thriller
マトリックス	Action Sci-Fi Thriller
スライダター	Action Sci-Fi Thriller
インデペンデンス・デイ	Action Sci-Fi War
スターゲイト	Action Adventure Sci-Fi
ファイブ・エレメント	Action Sci-Fi
Star Wars: Episode VI - Return of the Jedi	Action Adventure Romance Sci-Fi War
エイリアン	Action Horror Sci-Fi Thriller
僕の恋星	Action Sci-Fi
カタカ	Drama Sci-Fi Thriller
博士の異常な愛情または私は如何にして心	Sci-Fi War
ニューヨーク1997	Action Adventure Sci-Fi Thriller
今そこにある危機	Action Adventure Thriller
スタートレック ファーストエンタクト	Action Adventure Sci-Fi
ターミネーター	Action Sci-Fi Thriller
スターウォーズ エピソード1 ファントム・メ	Action Adventure Fantasy Sci-Fi
ロスト・ワールド/ジュラシックパーク	Action Adventure Sci-Fi Thriller
トゥモロー・ネバー・ダイ	Action Romance Thriller
k200のときの推薦結果	
ターミネーター2	Action Sci-Fi Thriller
マミー	Action Adventure Horror Thriller
スライダター	Action Sci-Fi Thriller
トータル・リコール	Action Adventure Sci-Fi Thriller
僕の恋星	Action Sci-Fi
ターミネーター	Action Sci-Fi Thriller
スターゲイト	Action Adventure Sci-Fi
ファイブ・エレメント	Action Sci-Fi
ニューヨーク1997	Action Adventure Sci-Fi Thriller
博士の異常な愛情または私は如何にして心	Sci-Fi War
今そこにある危機	Action Adventure Thriller
カタカ	Drama Sci-Fi Thriller
スタートレック ファーストエンタクト	Action Adventure Sci-Fi
群れをこ	Action Adventure Fantasy Sci-Fi
スタートレック0 未知への世界	Action Adventure Sci-Fi
エイリアン	Action Horror Sci-Fi Thriller
宇宙戦争	Action Sci-Fi War
トゥモロー・ネバー・ダイ	Action Romance Thriller
スタートレック2 ミスター・スポックを捜せ	Action Adventure Sci-Fi
007/ワールド・イズ・ノット・イナフ	Action Thriller
k300のときの推薦結果	
秋のソナタ	Drama
ロード・オブ・アライズーション	Horror
FLIRT フラット	Drama
ロスト・イン・スペース	Action Sci-Fi Thriller
Palmetto (邦名なし)	Film-Noir Mystery Thriller
Gas Bottom Boat The (邦名なし)	Comedy Romance
ぼくの嵐 パパの嵐	Comedy
Mummy's Hand The (邦名なし)	Horror
相棒人	Drama Thriller
ブレイクファスト・クラブ	Comedy Drama
ヤングガン2	Action Comedy Western
ソックスオン	Drama
大人は戦わなくていい	Drama
ラブ・オブ・ザ・ゲーム	Comedy Drama
カレンジャー・ガール	Drama
Eyes of Laura Mars (邦名なし)	Mystery Thriller
Defying Gravity (不明)	Drama
ペイルライダー	Western
バッシュマン・フィッシュ	Drama
スカーレット・レター	Drama
k400のときの推薦結果	
持ちきれなくて...	Comedy Drama Romance
02 Pick-Up (邦名なし)	Action Mystery Thriller
母がシェーンに恋したかい?	Drama Thriller
ザ・ターゲット	Thriller
Impact (不明)	Crime Drama
グッドナイト・ムーン	Drama
地球の危機	Adventure Sci-Fi
アメリカン・バップアロー	Drama
ドマーゴ	Drama Musical
ミセス・パーカー〜ジャズエイジの妻〜	Drama
卒業の輝き	Drama
サブリー・メールマン	Comedy
Giant Gila Monster The (邦名なし)	Horror Sci-Fi
ゴールド・ディカーズー伝説の秘宝を盗みこ	Adventure Children's
アメリカン・プレジデント	Comedy Drama Romance
ドック・パーク	Comedy Romance
エウジニスト	Horror
クルーエル・インテンションズ	Drama
グッドフォローズ	Crime Drama
プロロー宇宙からの不明物体ー	Horror Sci-Fi

表6-6を見ると、 $k=10$ 、 $k=20$ のときは推薦アイテムの90%以上がaction映画である。ところが、 $k=30$ となったとたんに、action映画はほとんど推薦されなくなり、また、 $k=30$ のときと $k=40$ のときの提示結果に重複している映画がまったく見られない。この事象は、どっちつかずの判断をした対による影響によるものと考えてよいだろう。予備実験に協力してくれた被験者からも、判断が難しく中点付近を評価することケースが多々あったという意見があった。

以上のような検証から、 k の設定について以下のような知見を得ることができた。

- ・ 推薦結果は k の値に大きく依存せず、むしろ p の値によるところが大きい。目的は活動利用者の入力コスト削減にあるので、 k の値は10から20程度でよいだろう。
- ・ p の値が0.5付近になる対が多いと、結果は理想から大きく外れてしまう。そのような場合、利用者に積極的な理由があるのであれば尊重すべきであるが、判断できずに中点付近を指定したのであれば考慮からはずすことを検討すべきである。

第7章

本実験とその評価・考察

本章では、提案手法による推薦と既存システムによる推薦との比較実験を行った結果について述べる。比較対象として、映画推薦システムMovieLens[5]を用いた。

7. 1 実験条件

本実験でも、所属研究室の学生数名に協力してもらい、提案システムの検証を行った。予備実験から得られた知見と、実験対象との条件を整えることを考慮し、実験条件を以下のように設定した。

- ・ 各個体の位置情報は、特異値分解による解析により得られた座標を使用。
- ・ 位置情報は上位1000軸を使用 ($n = 1000$)。
- ・ 既存システムに対する活動利用者の投票アイテム数、および提案システムに投票する一対比較評価数は15に設定した ($k = 15$)。

7. 2 実験内容

提案システムの実験手順は予備実験とほぼ同様に行われた。ただし、予備実験から得られた知見を踏まえ、システムが提示する映画の組み合わせについて被験者は必ずしも評価せずともよい、というオプションを加えた。これにより、被験者は判断の難しい対の評価を避けることができる。そして、一対比較数が15に達すると同時に評価を終了させた。推薦する映画数は20に設定した。

比較対象とした既存システムMovieLensの処理の流れについて述べる。その手順は以下のとおりである。

1. 被験者は、図3-2のようなシステムが提示する映画のリストの中から、見たことのある映画を選択し、5段階評価を行う。
2. システムが推薦する映画に対して5段階評価を行う。

こうして得られる両システムの推薦結果を推薦精度の比較に用いた。また、実験終了後の被験者に対してアンケートをとった。そして、アンケートの結果を、3.3節で述べた本研究の目的が達せられたかどうかの判断基準とした。

7. 3 結果とその検証

7. 3. 1 推薦精度の検証

被験者が評価した、システムごとの平均スコアの一部を表6-7に掲載する。

表6-7. システムごとの平均スコア

システム ユーザー	提案手法	MovieLens
A	3.80	3.92
B	3.07	4.00
C	2.83	2.50
D	3.67	2.85
E	2.93	2.50

本実験で得られたスコアに基づき、被験者ごとにwelchのt検定で分析したところ、有意水準5%で、「標本ごとの平均値に差がない」という仮説を棄却することができたのは被験者Dの標本のみであった。しかも、被験者Dは提案手法による推薦結果に高い評価をしている。以上より、提案手法による推薦精度が既存システムのそれと比較して、それほど遜色ないことが証明できた。

7. 3. 2 入力コストの検証

アンケートでは以下のような項目に答えてもらった。すべて自由記述で回答してもらった。

1. 入力コストの面で改善があったと感じたかどうか。（処理時間などの統計的な側面よりもむしろ、疲労度・モチベーションが保持できたか、などの心理的な側面から回答）
2. 一対比較法とSD法のどちらが評価しやすかったか
3. 自身の「こんな映画が見たい」という希望が、推薦結果に反映されているといえるかどうか

まず、「入力コスト」の比較をする上で、統計的尺度となりうる「処理時間」について重視しなかった理由について述べる。MovieLensでは、システムが提示するリストの中から利用者が見たことのある映画を評価していくため、「既知の映画を探すプロセス」が利用者の負担となっていることについては既に述べた。その上、MovieLensはすべて英語で実装されているため、日本人が使用する場合、「英語の出来不出来」「英語タイトルと日本語タイトルとは必ずしも一致しない」という要素がさらに負担を増大させている。MovieLensのアルゴリズムを本研究において日本語で実装できればよかったのだが、「映画をどういう順序で提示すれば最適化を測れるか」について言及した論文は存在するものの[3]、現存のMovieLensの映画整列アルゴリズムが載った論文は存在しない。以上のような理由から、「処理時間」を用いた入力コストの面での比較は、公平な条件下で比較することができないため、本研究では参考程度に紹介する。

質問1の回答について言及する。まず、「処理時間」についてだが、MovieLensで15タイトル評価するのに確認した総映画数は、少ない人で60から70、多い人で約1000にも達した。それに対して、提案手法では全員が、提示された50対の組み合わせから、評価可能な15の対を探し出すことができた。普段あまり映画を見ない利用者にとっては、提案手法のほうを使いやすいといえるかもしれない。また、質問に対して、「疲労度やモチベーションの面で改善が図れた」と答えた被験者は全体の6割程度であった。残りの4割は「どちらともいえない」という見解だった。提案手法に好意的な意見としては、

- ・ 映画同士を比較する楽しさがあり、モチベーションを保てた。
- ・ 評価基準が明確なので、結果に反映されそうな期待感がある。
- ・ 見たい映画の特徴を想定しながら、評価することができる。

といったものがあつた。否定的な意見としては、

- ・ 対を評価するかしないかの判断が既存システムに比べて難しい。

といった意見があつた。

次に、質問2について検証する。提案手法のほうの評価しやすいと答えた被験者は、全体の8割程度であった。提案手法に好意的な意見としては、

- ・ 提案手法は相対評価であるため、評価がしやすい。
- ・ どちらが好きかを選択するだけでよいので判断が容易。

といったものがあつた。否定的な意見として、

・嗜好比率の設定に整合性を設けるのが困難

という意見があったが、提案手法を支持する声が多かった。

最後に質問3についてだが、賛否両論が五分五分であった。推薦がうまくいかなかったケースについて考察する。

提案手法は、「スコアのつけ方に個人差がある」という点で、従来の手法から改良が施されていない。そこで、うまくいかなかった原因はスコアの個人差にあると考える。例えば、提示された対に対して、「左側の映画が右側の映画より結構好きだ」という利用者は、スコア p を0.1から0.3ぐらいの値に設定するだろう。そこで、被験者が評価したスコア p をそれぞれ0.1ずつシフトさせてみて、推薦結果にどう影響するか検証してみる。

表6-8. 値 p が推薦結果に与える影響

シフト前	シフト後
Borrowers The (邦名なし)	Overnight Delivery (邦名なし)
Hype! (邦名なし)	Mr. Magoo (邦名なし アメリカTVシリーズ?)
All the Vermeers in New York (邦名なし)	ディック・トレイシー
大脱走	悪い種子
Program The (邦名なし)	フレンチ・キス
ブレードランナー	Heaven's Burning (邦名なし)
母の眠り	エリック・ロメール 恋の秋
エントラップメント	007/ロシアより愛をこめて
ブロークダウン・パレス	リアリティ・バイツ
バッド・ガールズ	Two if by Sea(邦名なし)
Saltmen of Tibet The (邦名なし)	間諜最後の日
ワイルド・マン・ブルース	Head On (邦名なし)
街の灯	オードリー・ローズ/家
ジャイアント・ビーチ	ロウヘッド・レックス
子熊物語	クライム アンド パニッシュメント
Judy Berlin (邦名なし)	ホーリーマン
Inheritors The (Die Siebtelbauern) (邦名なし)	ハード・ターゲット
追撃者	デューン/砂の惑星
HANA-BI	ある女の存在証明
シクロ	Marlene Dietrich: Shadow and Light (不明)

表 6 - 8 の左側の列はある被験者に提示された推薦結果、右側の列はその被験者のスコア全てを0.1ずつ「好きな」映画のほうにシフトさせたときに提示される推薦結果である。評価対を増やしていったときと同じように、結果がまるで重複していないことがわかる。以上のことから、提案手法はある程度利用者の嗜好を特定することはできるものの、 p の値が推薦結果に与える影響が大きいことがわかる。

次に比較対象を用意する必要性について言及する。結果に自身の意思が反映されていないとした被験者の全てが、「推薦結果に嫌いな要素を含む映画が多く含まれていた」と回答した。その理由は明らかであり、評価する映画を被験者に自由に選択させたという点にある。つまり、このような被験者は自身の嫌いな要素を含む映画について評価をしていないのである。こうなると、MovieLensのような既存システムと同じようにアイテムをランダムに評価してもらう必要があると思われるかもしれないがそうではない。提案手法はそのアルゴリズム上の構造から、嫌いな要素を含む映画を2～3タイトル追加するだけで十分だと考えられる。

第8章

結論と今後の課題

本研究の最大の目的として「活動利用者の入力コストの軽減」を掲げた。その上で、「推薦精度を落とさない」こと、「利用者が今見たいアイテムを推薦するシステム」の実現もあわせて目的とした。本研究では、それぞれの項目について以下のように結論付ける。

8. 1 推薦精度について

活動利用者の嗜好情報入力が少ない場合において、本研究の提案手法による推薦精度は、既存の推薦システムの推薦精度と比較し、それほど遜色ない数値を示した。しかし、MovieLensのような協調フィルタリング技術を用いた推薦システムは、本来、利用者による入力が多ければ多いほど、推薦精度が洗練されていくシステムである。一方、提案手法に対して、ある程度以上、一対比較入力数を増やしたとしても精度に大きな効果が現れるとは思えない。このことから、提案手法は、面倒な処理を嫌がり、すぐに結果をほしがるような利用者に対しては有効な手段となりうる。

8. 2 入力コストについて

本研究では、一対比較法によって活動利用者の嗜好を表現する手法を提案した。従来のSD法による表現手段と比較して、

- ・ 活動利用者は絶対評価を行うことが可能となり、判断の基準が明確になった。
- ・ 従来のシステムほど、「嫌いなコンテンツを評価させなければならない」という点で、固執する必要がない。
- ・ 一対比較による評価に「楽しさ」が生まれ、システムの利用にモチベーションを保つことができる。

といった点で、入力コストを軽減させることができたといえる。

一方で、

- ・ SD法と比較し、評価数が増えていくにつれて、整合性を保つことが難しくなる。

という課題を残した。

8. 3 今後の課題

8. 3. 1 使用した履歴データについて

本研究の実験で用いたアイテムの位置情報は、既存システムの履歴データ、つまりSD法による評価が施されたデータから得たものである。これは、推薦システムの検証を行うためには大量のデータが必要であり、個人の力では収集することが不可能であるという点が原因である。提案手法の検証には本来、一対比較法により収集したデータから解析したアイテムの位置情報を用いる必要があると思われる。

8.3.2 嗜好空間形成アルゴリズムについて

提案手法では、一対比較値 p により、お互いのアイテムの位置情報を $p:(1-p)$ に分割する超平面をひとつの評価から得られる利用者の嗜好空間（図8-1の赤線）であると定義した。この定義は「お互いの映画がもつ全ての属性に対して、好きだと評価した映画が持つ属性が利用者に合う属性である」という仮定に基づくものである。例として、利用者がアイテムAとアイテムBを比較し、 $p=0.5$ としたときについて考える。このとき、利用者があるひとつの属性に基づき判断を下したのであれば、提案手法ではほぼ問題はないと考えられる。なぜなら、そのほかの属性に関して、2つのアイテム間に差異はほとんどないと考えて差し支えないからである。しかし、利用者が「第1軸の属性に基づけばBのほうが嗜好に合い、第2軸の属性に基づけばAのほうが嗜好に合う。総合的に判断して同等である。」という考え方をしたとするなら、提案手法による嗜好空間の形成に矛盾が生じる。図8-1でいえば、緑で表現した超平面を形成すべきである。（この考え方は、AHPのアルゴリズムそのものである）

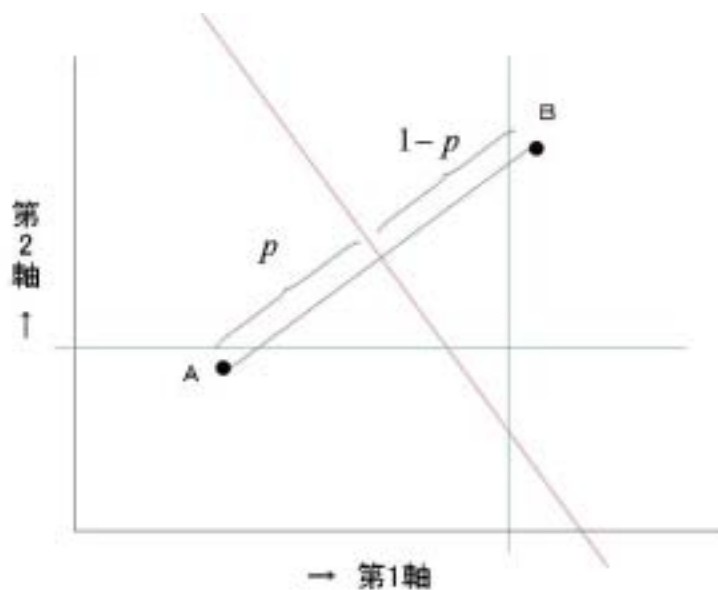


図8-1. 嗜好空間

以上のような理由から、一対比較による評価が複数の属性に基づく評価であった場合、提案手法では、正確な嗜好空間の形成が阻害されるだろう。

8. 3. 3 推薦アルゴリズムについて

本文中で、ある要因を属性として持つアイテムに対して一対比較の評価を極端に寄せると、その要因を持つアイテムが推薦される可能性が高い、ということを述べた。しかし、このような条件を下とする要求は、好むアイテムを指定して「そのアイテムに最も近いアイテムを教えてください」という要求をしているのとほとんど変わらないといえる。言い換えれば、人気のないアイテムも網羅的に推薦しようとするのが精度に影響しているといえる。今後、多くの利用者に注目されないアイテムを推薦できるアルゴリズムの開発が求められる。

参考文献

- [1] Breese, J. S., Heckerman, D., & Kadie, C.:
“Empirical Analysis of Predictive Algorithms for Collaborative Filtering”,
MSR-TR-98-12, 1998
- [2] 徳永健伸：
情報検索と言語処理，
東京大学出版会，1999
- [3] <http://www.amazon.com/gp/homepage.html/>
- [4] Al Mamunur Rashid, Istvan Albert, Dan Cosley, Shyong K.Lam, Sean M.McNee,
Joseph A.Konstan, John Riedl：
“Learning New User Preferences in Recommender Systems”，
In Proceedings of the International Conference on Intelligent User Interfaces, pp. 127-134
，2002
- [5] <http://movielens.umn.edu/>
- [6] 神寫敏弘：
なんとなく協調フィルタリング —順序応答に基づく推薦，
The 17th Annual Conference of the Japanese Society for Artificial Intelligence，2004
- [7] 中森義輝：
感性データ解析—感性情報処理のためのファジィ数量分析手法，
森北出版, 2000
- [8] <http://jamesthornton.com/cf/>
- [9] 堀田政二：
クラスタリングに基づく情報の検索と視覚化，
九州芸術工科大学博士論文，2002
- [10] R. Xiong， M. A. Smith and S. M. Drucker:
“Visualization of collaborative information for end-users”，
MSR-TR-98-52，1998
- [11] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl：
“Item-based Collaborative Filtering Recommendation Algorithms”，
In Proc of the 10th International World Wide Web Conference，Hong Kong，2001
- [12] 林知己夫：
数量化—理論と方法—，
朝倉書店，1993
- [13] 加藤豊，小沢正典：

ORの基礎 AHPから最適化まで ,
実教出版 , 1998

[14] <http://www.interscope.co.jp/method/m03.html>

[15] 神寫敏弘:

なんとなく協調フィルタリング – 複数の順序応答に基づく推薦 ,

The 19th Annual Conference of the Japanese Society for Artificial Intelligence , 2005

[16] 岩下豊彦 :

SD法によるイメージの測定 ,

川島書店 , 1983