

2015 (平成 27) 年度 修士論文

**価値観アイテムモデリングを用いた
推薦システムに関する研究**

**Study on Recommender System
Using Personal-value-based Item Modeling**



首都大学東京大学院 システムデザイン研究科 システムデザイン専攻
情報通信システム学域 高間研究室

学修番号：14890535 氏名：山口 貴之
指導教員：高間 康史 教授

目次

1	はじめに	1
2	関連研究	3
2.1	代表的な推薦アルゴリズム	3
2.1.1	内容ベースフィルタリング	3
2.1.2	協調フィルタリング	5
2.2	価値観モデリングを利用した研究	9
2.2.1	価値観に着目した研究	9
2.2.2	価値観ユーザモデリングに関する研究	10
2.3	推薦説明に関する研究	14
2.3.1	Transparency (透明性)	14
2.3.2	Scrutability (理解度)	15
2.3.3	Trust (信頼性)	16
2.3.4	Persuasiveness (説得性)	16
2.3.5	Effectiveness (有効性)	16
2.3.6	Efficiency (効率性)	17
2.3.7	Satisfaction (満足度)	17
3	価値観アイテムモデリング手法	18
3.1	アイテムモデリングの利用	18
3.2	RMRate の拡張	21
3.2.1	相関ルールについて	21
3.2.2	リフト値	22
3.2.3	リフト値によるモデリング方法	24
3.3	アイテムモデルのアイテムベースによる推薦への適用	26
3.3.1	リフト値の類似度計算への適用方法	26
3.3.2	アイテム間類似度の計算方法	27
3.4	アイテムモデルの推薦説明への適用	29
4	評価実験	32
4.1	アイテムモデルの解釈に関する予備実験	32
4.1.1	実験概要	32
4.1.2	実験結果	39
4.1.3	推薦システムでの利用に関する考察	44
4.2	評価実験	46
4.2.1	実験概要	46

4.2.2	システムで利用するデータセット	48
4.2.3	システムにおける推薦説明の提示方法	50
4.2.4	システムで利用する推薦アルゴリズム	54
4.2.5	プロトタイプシステムとデータの収集方法	57
4.2.6	評価実験の結果と考察	62
5	おわりに	72
	謝辞	73
	参考文献	74
	発表文献	79

1 はじめに

近年、情報化技術の発展により、人の扱う情報量はますます多くなってきている。ショッピングサイトの Amazon¹ では1億種類以上の商品数を扱い、動画共有サイトの Youtube² では1分間に合計300時間を超える動画がアップロードされている。このような、膨大な情報の中からユーザが自分のニーズに合ったものを探すのは困難であり、情報過多 (information overflow) [1] と呼ばれ、問題視されている。これに対する解決策の一つとして、ユーザの行動履歴をもとに情報のフィルタリングを行い有用性の高いものを推薦する、情報推薦システムが注目されている。情報推薦の代表的な手法に協調フィルタリングがあり、ユーザの行動履歴と類似する履歴を持つユーザの評価を利用して推薦アイテムを決定する。この手法は Amazon を始めとするショッピングサイトを含め、幅広く使用されている。

情報推薦の手法の中にはユーザの嗜好だけでなく、意思決定の背後にある価値観に着目して推薦を行う手法が提案されている [2]。価値観とはユーザがアイテムのどの要素を重視するかというものであり、消費行動に影響を与える要素であることからマーケティング等に利用されている。文献 [2] で提案されている価値観モデリング手法では、価値観をアイテムの各属性に対するユーザのこだわりとみなし、その強さを評価一致率 RMRate (Rating Matching Rate) と呼ばれる指標によって算出することで、ユーザレビューからユーザモデルを構築している。ユーザベースの協調フィルタリングや内容ベースの推薦手法に適用し、その有効性が示されている [2, 3]。

本論文ではこの手法をアイテムモデリングに拡張することを目的とする。アイテムに投稿されたレビューを収集し、同様の分析を行うことで、そのアイテムがどの属性に着目して評価されているのかを価値観モデルとして求めることが可能と考える。価値観モデル構築に利用可能なレビューは、アイテムの方がユーザモデリングと比較して集めやすいことが期待できる。また、アイテム間の類似度計算に構築されたアイテムモデルを使用することで、アイテムベースの協調フィルタリング [15] への適用なども可能となる。さらに、構築されたアイテムモデルをアイテムの特徴として捉えユーザに提示することで、推薦説明に活用する事も期待できる。情報推薦において、推薦アイテムと共にその推薦理由やアイテムの特徴を説明することは、推薦システムに対する信頼性や満足度を向上させる手段として、近年その必要性が高まってきている [35]。推薦説明に関する研究では、対象ユーザの行動履歴や対象ユーザと類似する他のユーザの情報に基づいて説明の提示を行う手法が一般的であるが、価値観モデルに基づく推薦説明は、従来手法とは異なる観点からの情報提供に繋がることが期待できる。

¹<http://www.amazon.com>

²<http://www.youtube.com>

本論文では価値観に基づくアイテムモデリング手法を提案する。また構築されたモデルを用いた、推薦アルゴリズム、推薦時のアイテム説明の生成手法も提案する。アイテムモデルはモデル構築に多数のレビューを利用可能という利点から、RMRateの代わりに相関ルールの指標の1つであるリフト値 [5] を算出し、アイテムの特性をより詳しく分析する。具体的には、アイテムに対する好評・不評、属性に対する好評・不評の組み合わせ4種類に分類してリフト値を算出する。推薦アルゴリズムでは、アイテム間の類似度計算に提案するアイテムモデルを使用しアイテムベースの推薦を行う。さらに、構築されたアイテムモデルから説明文を作成、可視化することで推薦時にユーザに提示する。

提案手法の有効性を検証するため、プロトタイプシステムを構築して評価実験を行う。また、アイテムモデルの解釈に関する予備実験を事前に行う。予備実験ではいくつかのアイテムモデルに対するユーザの解釈をアンケートにより収集する。その結果に基づき、提案するアイテムモデルを利用した推薦説明の生成方法について検討する。評価実験では推薦アルゴリズムと推薦説明の有効性の観点から検証を行う。推薦アルゴリズムでは、提案手法、および比較手法としてアイテムベースの協調フィルタリングによる推薦を行い、精度を比較する。推薦説明に関しては、提案手法で作成した推薦説明をアイテムと同時に提示し、それがアイテムを評価する際に参考になったかどうかを回答してもらう。また評価実験後にアンケートを行うことでシステム全体に対する意見も収集する。実験結果より提案手法の推薦精度は従来手法と同等以上であり、推薦説明も有用である事を示す。

2 関連研究

本章では情報推薦における代表的な推薦アルゴリズムを述べた後に、価値観モデリングを利用した研究、推薦説明の研究について説明する。

2.1 代表的な推薦アルゴリズム

2.1.1 内容ベースフィルタリング

情報推薦には様々な手法が存在し、その中でも代表的なものに内容ベースフィルタリングと協調フィルタリングがある。内容ベースフィルタリングとはユーザーのプロファイルを作成し、それにあった特徴を含むアイテムを推薦する手法である [7, 8]。プロファイルはアイテムの持つ属性値により構成され、ユーザーの検索や評価履歴などから作成される。図 2.1 にその概要と例を示す。図 2.1 の場合、例えば映画であれば、好きなジャンル、監督、俳優などが属性値であり、プロファイルを構成する。推薦時には、これらの属性値、あるいは関連の高い属性値を持つアイテムを推薦する。このようなアイテムの持つメタデータ以外にも、文章の持つ特徴量を利用した推薦も存在する [8]。特徴抽出には主に TF-IDF [11] と呼ばれる指標が使われ、その文章における各単語の重要度を値として持つベクトルとして表現する。これにより、ユーザーの閲覧した文章（ニュース記事など）を元にプロファイルを作成したり、プロファイルに近い特徴量を持つ文章をユーザーに推薦する。

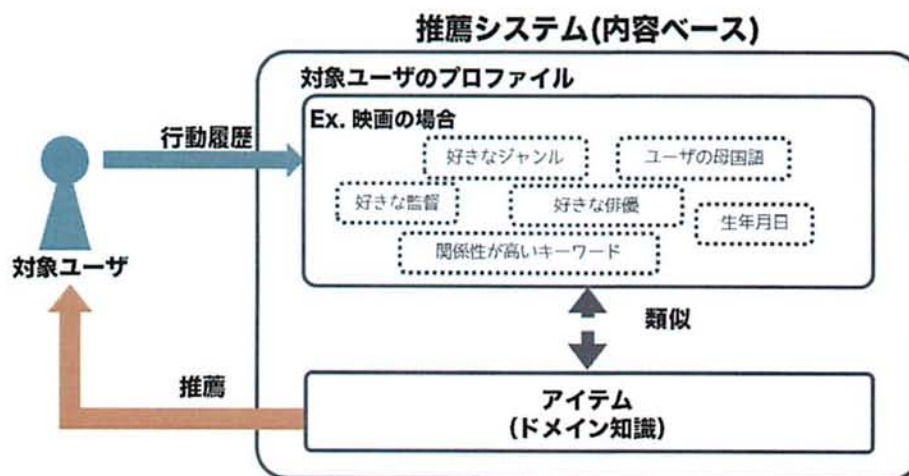


図 2.1: 内容ベースフィルタリングの仕組み

内容ベースフィルタリングでは、ユーザからの評価が少数しか得られていない新規アイテムや人気のないアイテムを推薦できる利点がある。一方、ユーザの要求を明示的に得ることが難しいこと、アイテムのプロファイルを作成するのにドメイン知識が必要であることといった課題があるため、実際のシステムに実装することは難しいとされている [10]。またユーザプロフィールに近いアイテムばかり推薦されるので、推薦されるアイテムの多様性に欠けることも指摘されている [9, 10]。



2.1.2 協調フィルタリング

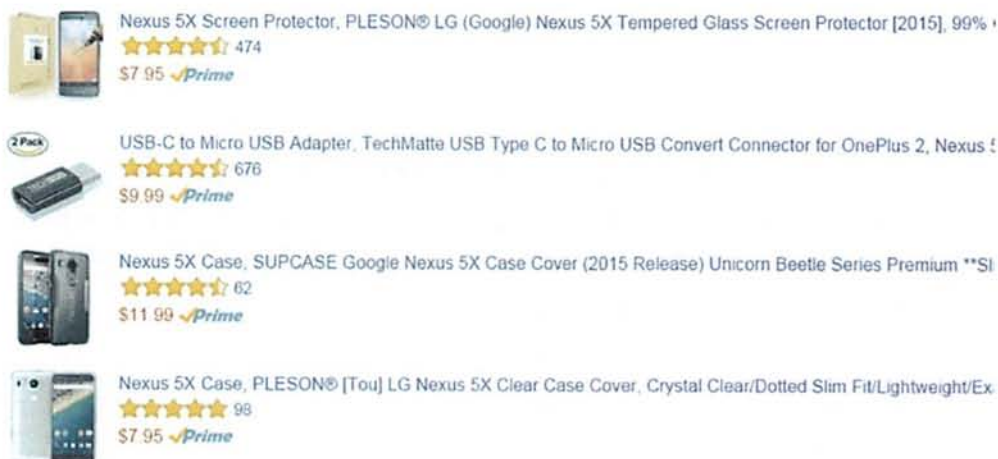
協調フィルタリングとは、アイテムの持つ属性や属性値といったドメイン知識を利用せず、他人の口コミを利用した推薦方法であり [12]，実装が容易であること，内容ベースの推薦と比べて多様性のある推薦が可能であることなどの利点から，Amazonをはじめとするショッピングサイト等で幅広く利用されている．図 2.2 は実際に協調フィルタリングを用いた Amazon における推薦の例であり，他ユーザのアイテムへの評価履歴を元にアイテムを推薦している．

Customers Who Bought This Item Also Bought



Item	Price	Rating	Reviews
Nexus 5X Screen Protector, PLESON® LG (Google) Nexus 5X Tempered Glass Screen...	\$7.95 ✓Prime	★★★★★	474
Nexus 5X Case, Spigen® [Thin Fit] Exact-Fit [Black] Premium Matte Finish Hard Case for Nexus 5X...	\$11.99 ✓Prime	★★★★★	172
USB-C to Micro USB Adapter, TechMatte USB Type C to Micro USB Convert Connector for...	\$9.99 ✓Prime	★★★★★	676
Type C Cable, Yoozon® 3.3ft/1m USB 2.0 Type C Type A (USB-C to USB-A Cable for Nexus...	\$7.99 ✓Prime	★★★★★	620

What Other Items Do Customers Buy After Viewing This Item?



Item	Price	Rating	Reviews
Nexus 5X Screen Protector, PLESON® LG (Google) Nexus 5X Tempered Glass Screen Protector [2015], 99%+	\$7.95 ✓Prime	★★★★★	474
USB-C to Micro USB Adapter, TechMatte USB Type C to Micro USB Convert Connector for OnePlus 2, Nexus 5...	\$9.99 ✓Prime	★★★★★	676
Nexus 5X Case, SUPCASE Google Nexus 5X Case Cover (2015 Release) Unicorn Beetle Series Premium **SI	\$11.99 ✓Prime	★★★★★	62
Nexus 5X Case, PLESON® [Tou] LG Nexus 5X Clear Case Cover, Crystal Clear/Dotted Slim Fit/Lightweight/Ex	\$7.95 ✓Prime	★★★★★	98

図 2.2: Amazon における協調フィルタリングの使用例

協調フィルタリングの代表的な手法にはユーザベース手法 [14] があり，ユーザ同士の嗜好の類似性に基づいて推薦を行う．図 2.3 にユーザベースによる協調フィルタリングの仕組みを示す．図 2.3 のように，ユーザベースでは対象ユーザの行動履歴を元に類似ユーザを予測し，その類似ユーザが好むアイテムを対象ユーザに推薦する．代表的な例として映画推薦サイトの MovieLens が挙げられる．

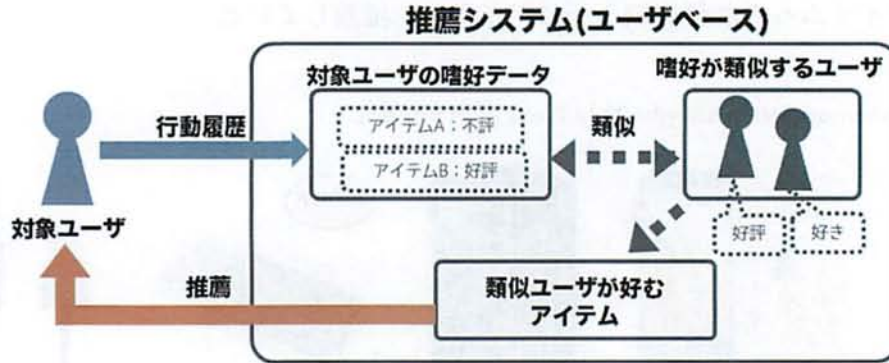


図 2.3: ユーザベースの協調フィルタリングの仕組み

ユーザベースのアルゴリズムを MovieLens で提案されたものに基づいて説明する．ユーザベースでは推薦を以下の2段階で行う．

- (1) 嗜好の類似度計算：対象ユーザのアイテム評価や購入履歴より，他ユーザとの嗜好の類似度を計算
- (2) 評価値の予測：(1)の類似度を利用し，対象ユーザの未評価アイテムへの予測評価値を計算

(1)の行程について説明する．ユーザベースの推薦においては，ユーザが投稿した評価値を元にユークリッド距離やコサイン距離，または Pearson 相関係数を算出し，ユーザ間の類似度として扱うことが多い [7]．その中でも Pearson 相関を用いた場合，良い結果が出ることが報告されている [17]．以下に Pearson 相関を利用した場合のアイテム i_a と i_b の類似度 $sim_{item}(i_a, i_b)$ の計算式を示す．

全ユーザの集合を $U = \{u_1, u_2, \dots, u_n\}$ ，全アイテム集合を $I = \{i_1, i_2, \dots, i_m\}$ とした時，Pearson 相関を用いた，ユーザ u_x とユーザ u_y の類似度 $sim_{user}(u_x, u_y)$ は式 2.1 により定義される．

$$sim_{user}(u_x, u_y) = \frac{\sum_{i_k \in I_{xy}} (r_{xk} - \bar{r}_x)(r_{yk} - \bar{r}_y)}{\sqrt{\sum_{i_k \in I_{xy}} (r_{xk} - \bar{r}_x)^2} \sqrt{\sum_{i_k \in I_{xy}} (r_{yk} - \bar{r}_y)^2}} \quad (2.1)$$

ここで I_{xy} は u_x, u_y が共に投稿したアイテムの集合, r_{xk} は u_x のアイテム i_k に対する評価値, \bar{r}_x は u_x が投稿した全アイテムに対する評価値の平均値を示す.

(2) の行程では, 推薦対象ユーザ u_t のアイテムへの評価値を予測する. u_t のある未評価アイテム i_a への予測評価値 \hat{S}_{ta} は式 2.2 によって計算される. ここで, U_a は i_a を評価しているユーザの集合を示している. 予測評価値は u_t と u_x の類似度を重みとした, i_a に対する評価値の加重平均によって計算される. $|sim_{user}(u_t, u_x)|$ は $sim_{user}(u_t, u_x)$ の絶対値を意味する.

$$\hat{S}_{ta} = \bar{r}_t + \frac{\sum_{u_x \in U_a} sim_{user}(u_t, u_x)(r_{xa} - \bar{r}_x)}{\sum_{u_x \in U_a} |sim_{user}(u_t, u_x)|} \quad (2.2)$$

ユーザベースの推薦手法ではアイテムのドメイン知識を利用せず, 評価値のみで推薦を行うことができる. 一方で式 2.1 からわかるように, データセットの規模が大きくなるほど類似度の計算量が大きくなる問題が指摘されている [16]. 文献 [16] ではユーザをランダムサンプリングする, 購入量の少ないユーザを除く, 購入数が極端に多いまたは少ないアイテムを省く, などの手法をとることで計算量を減らす工夫をしているが, 推薦の品質, 精度が低下すると報告している.

協調フィルタリングの手法にはユーザベースの他にアイテムベースの推薦手法 [15,16] が存在する. アイテムベースの概要を図 2.4 に示す. ユーザベースではユー

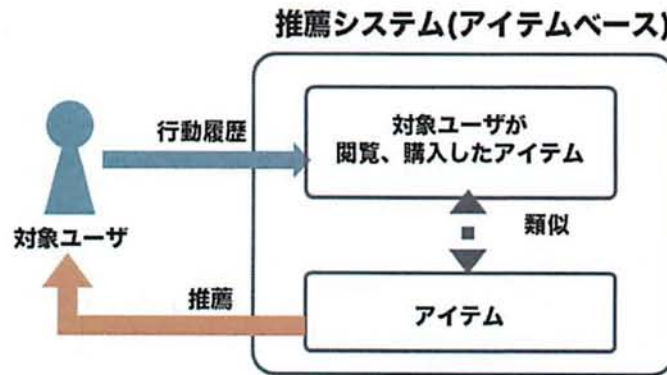


図 2.4: アイテムベースの協調フィルタリングの仕組み

ザ間の類似度を計算するのに対し, アイテムベースでは図 2.4 のように, アイテム間の類似度を計算し, 対象ユーザが選択した (または好む) アイテムと類似するアイテムを推薦する. アイテムベースを使用した代表的なサイトとして Amazon が挙げられる [16].

アイテムベースの推薦においては, アイテムに投稿されている評価値を元にユークリッド距離やコサイン距離, または Pearson 相関係数を算出し, それをアイテム

間の類似度として扱うことが多い [16]. 以下に Pearson 相関を利用した場合のアイテム i_a と i_b の類似度 $sim_{item}(i_a, i_b)$ の計算式を示す.

$$sim_{item}(i_a, i_b) = \frac{\sum_{u_x \in U_{ab}} (r_{xa} - \bar{r}_a)(r_{xb} - \bar{r}_b)}{\sqrt{\sum_{u_x \in U_{ab}} (r_{xa} - \bar{r}_a)^2} \sqrt{\sum_{u_x \in U_{ab}} (r_{xb} - \bar{r}_b)^2}} \quad (2.3)$$

ここで U_{ab} は i_a, i_b に共に評価したユーザの集合, r_{xa} は u_x のアイテム i_a に対する評価値, \bar{r}_a は i_a に投稿されている評価値の平均値を示す.

アイテムベースの代表的な推薦アルゴリズムとして, Sarwar ら [15] は, MovieLens で提案された手法と同様に, アイテム間の類似度から, 対象ユーザの未評価アイテムに対する予測評価値を求める方法を提案している. 対象ユーザ u_t のアイテム i_a に対する予測評価値 \hat{S}_{ta} は, アイテム間の類似度を重みとして式 2.4 のように計算される. I_t は, u_t が過去に評価したアイテム集合を示し, r_{ik} は u_t が i_k に下した評価値, $sim_{item}(i_k, i_a)$ はアイテム i_k と i_a の類似度, $|sim_{item}(i_k, i_a)|$ は $sim_{item}(i_k, i_a)$ の絶対値を意味する.

$$\hat{S}_{ta} = \frac{\sum_{i_k \in I_t} r_{ik} sim_{item}(i_k, i_a)}{\sum_{i_k \in I_t} |sim_{item}(i_k, i_a)|} \quad (2.4)$$

ユーザベースの推薦手法では, 対象ユーザがある程度の数のアイテムを評価しなければ, 他ユーザとの類似度を求めることが困難になり, 満足のいく推薦が行えない. 一方でアイテムベースの推薦手法では, 式 2.4 からわかるように, アイテム間の類似度が計算されていれば, 対象ユーザの評価アイテム数が少なくてもアイテムを推薦することができる.

またアイテムベースの手法ではユーザベースと同様, データセットの規模が大きくなるほど類似度の計算量が増す問題があるが, アイテム間の類似度を事前にオフラインで計算しておくことで, オンラインでの計算量を減らすことができる.

前述のユーザベースやアイテムベースの推薦手法は, 蓄積されたデータより, 類似するユーザ (アイテム) を推薦時に計算して求める手法であった. このような手法はメモリベースと呼ばれている. 協調フィルタリングには一方で, 事前に調べておいたデータをあらかじめ統計的にモデル化しておくことで推薦を行う, モデルベースと呼ばれる手法がある [13]. モデルベースではモデル作成にクラスタリング, 効用関数や評価値を予測する関数, 確率モデルを使う方法があり, 事前にモデルを作成することで推薦時の計算コストを低くできる利点がある. 一方でユーザやアイテムの集合が変化した場合, モデルの作り直しが必要であるため, 最適化されたモデルを適用した適切な推薦を行うのに時間がかかる [8].

2.2 価値観モデリングを利用した研究

2.2.1 価値観に着目した研究

価値観とはユーザがアイテムのどの要素を重視するかというものであり、消費行動に影響を与える要素であることからマーケティング等に利用されている。文献 [18] では消費者の嗜好を調査するのに、価値観を 18 の要素に分類した Rokeach Value Survey (RVS) と呼ばれる手法を提案しており、多くの調査で使用されている。Holbrook は消費行動に影響を与える価値観を 8 つの要素に分類し、その検証方法を提案している [19]。価値観と関係が深い考えにパーソナリティが存在する。パーソナリティは「個人の内面に由来する一貫した行動様式およびそのプロセス」と定義されており [26]、調査する方法としてビッグファイブ理論 [27] やエゴグラム [28] などがある。パーソナリティも価値観と同様に人の行動や嗜好に影響を与えるが、パーソナリティは先天的な性質で一貫して変化しにくいとされるのに対し、価値観は後天的な性質で経験や環境に影響されやすいとされる [29]。

情報推薦への応用に関しては、調査した限りではパーソナリティを利用した研究の方が、価値観を利用した研究よりも現状では多く存在する。Roshchina は対象ユーザのレビューよりパーソナリティのモデリングを行い、協調フィルタリングと組合せることで推薦を行っている [20]。Hu らは楽曲の推薦において、事前にユーザにアンケートを行うことでユーザのパーソナリティを作成している [21]。その結果、楽曲に詳しくない人が、パーソナリティを利用して推薦されたアイテムの方を、パーソナリティを利用せずに推薦したアイテムよりも気に入ることを示している。

パーソナリティはその先天的な性質により、情報推薦システムでの利用においては嗜好との関連が希薄であることが指摘されている [25]。そこで、より人の意思決定や嗜好に関係が近い価値観を情報推薦に利用した研究も存在する。

Felfernig ら [24] は価値観を利用したモデリング手法である多属性効用理論 [22,23] を利用してユーザモデルを作成し、推薦を行っている。多属性効用理論では、ユーザのアイテムに対する効用 (utility) を、アイテムの持つ属性値に対するスコアと、それぞれの属性に対する重要度から重み付けを行うことでモデリングを行っている。そのため多属性効用理論ではアイテムの属性の他に、ドメイン知識を必要とする。2.2.2 項で詳しく紹介するが、服部ら [2] は価値観をユーザの属性に対するこだわりとみなし、アイテムの属性とそれに対する評価値からユーザモデルを作成することで、多属性効用理論と比べてより少ない情報から価値観を推論し、推薦を行うことに成功している。

2.2.2 価値観ユーザモデリングに関する研究

Hattori ら [2] は価値観を「ユーザの属性に対するこだわり」と定義してモデリングを行っている。ここでのこだわりとは「ユーザがどの属性を重視してアイテムの評価を決定するか」を意味し、図 2.5 にその例を示す。図 2.5 のように、テレビにおいて画質に対して強いこだわりを持つユーザがいた場合、値段や消費電力の優劣よりも、解像度や画質の良し悪しがテレビの評価に影響を与えと考えられる。

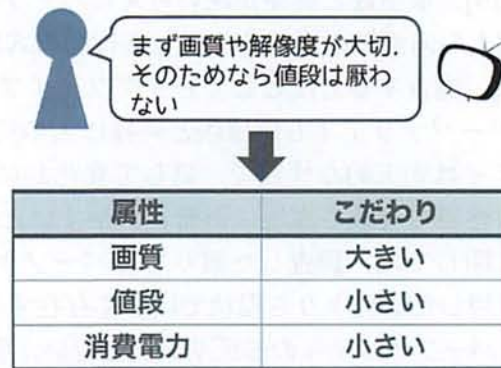


図 2.5: 文献 [2] における属性に対するこだわりの例

文献 [2] ではこのようなユーザのこだわりを、RMRate と呼ばれる指標を用いて算出し、ユーザモデルを作成している。RMRate は図 2.6 に示すようなレビューサイト (Yahoo!映画¹) のレビューを情報源として利用している。図 2.6 のようなレビューは、レビューの文章とアイテムに対する評価値以外に、各属性に対する評価値も含んでいる。価値観モデリングではこの属性の評価値を使うことでモデルを作成している。

以下に RMRate の計算方法を記述する。RMRate は、ユーザ u のアイテム i に対する評価極性 p_{ui} と、 i の属性 a に対する評価極性 p_{uia} から求められる。 p_{ui} と p_{uia} は、評価値の閾値による判定で求められる。極性判定の手順を説明する。始めに、 u が下した全アイテムの評価値の平均値 $\theta(u)$ を式 2.5 により計算する。

$$\theta(u) = \frac{1}{I} \sum_{i \in I} r_{ui} \quad (2.5)$$

ここで I はユーザ u が与えたアイテムの集合、 r_{ui} は u が i に投稿した評価値 (総合評価値) を意味する。 u が i の各属性に投稿した評価値 (属性評価値) の平均値 $\theta(u, i)$

¹<http://movies.yahoo.co.jp>



図 2.6: 属性評価がされているレビューの例 (Yahoo!映画)

を式 2.6 のように計算する.

$$\theta(u, i) = \frac{1}{|A_u|} \sum_{a \in A_u} r_{uia} \quad (2.6)$$

A_u はアイテムが持つ属性の内, ユーザ u が評価を与えた属性の集合, r_{uia} は u が i の属性 a に投稿した評価値を意味する. p_{ui} と p_{uia} は $\theta(u)$ と $\theta(u, i)$ をそれぞれ閾値とし, 式 2.7, 2.8 のように求められる.

$$p_{ui} = \begin{cases} pos. & (r_{ui} \geq \theta(u)) \\ neg. & (r_{ui} < \theta(u)) \end{cases} \quad (2.7)$$

$$p_{uia} = \begin{cases} pos. & (r_{uia} \geq \theta(u, i)) \\ neg. & (r_{uia} < \theta(u, i)) \end{cases} \quad (2.8)$$

u が評価した全アイテムの各属性 a について, p_{ui} と p_{uia} が一致した回数を $O(u, a)$, 不一致であった回数を $Q(u, a)$ とした場合, u における a に対する RMRate の値 P_{ua}

は式 2.9 のように計算される。RMRate は 0 から 1 までの値を持ち、値が大きいほどその属性に対するこだわりが強いとみなされる。

$$P_{ua} = \frac{O(u, a)}{O(u, a) + Q(u, a)} \quad (2.9)$$

ユーザレビューを元にした RMRate の計算例を図 2.7 に示す。図 2.7 ではユーザが複数のアイテムに対して評価を行った結果に基づき、各評価の極性判定を行なった後、属性毎に RMRate の算出をしている。RMRate の値より、この場合ユーザは画質に対してこだわりを持つことが推論できる。



図 2.7: RMRate の計算例

このように式 2.9 で求められた RMRate より、構築したユーザモデルは属性数を n とした n 次元ベクトルで表される。求めたベクトルより、式 2.1 と同様の Pearson 相関を用いて他のユーザとの類似度が計算され、ユーザの未評価のアイテムへの評価値を予測する。

データセットを用いたオフライン評価実験では、RMRate を協調フィルタリングと組み合わせて用いることで、低評価のアイテムにおける平均絶対誤差 MAE (Mean

Absolute Error) が低減する結果などが得られている [3]. 清水らは, レビューを利用したユーザモデルの構築方法に RMRate を使用することを提案し, 構築されたモデルを推薦に利用して, その有用性や特性を考察している [30]. 価値観モデリングを内容ベースフィルタリングに適用した推薦方法も存在する. 文献 [2] では RMRate が大きい属性について, それと同じ属性が高く評価されているアイテムを推薦したり, その属性が持つ属性値と類似する属性値を持つアイテムを推薦する手法を提案しており, ユーザ実験で有効性を示している.

表 2.1: 推薦説明の目指す推薦システムへの効果

目指す効果	効果の詳細
Transparency (透明性)	ユーザがアイテムの推薦過程, 理由を把握できる
Scrutability (理解度)	ユーザがその推薦が本当に正確かどうか判断できる
Trust (信頼性)	ユーザがシステムを信頼できる
Persuasiveness (説得性)	ユーザがアイテムを購入する (気に入る)
Effectiveness (有効性)	ユーザの意思決定, 判断の支援
Efficiency (効率性)	ユーザの意思決定を早める
Satisfaction (満足度)	システムの使いやすさ, 満足度を向上する

2.3 推薦説明に関する研究

情報推薦の評価は一般的に, 精度や再現率, 予測評価値と実際の値の差 (MAE) などで検証されるが, これらの指標だけでは推薦システムを十分に評価できていないという認識が高まってきている [31, 32]. 近年では推薦アイテムの多様性 [33], 新しさや意外性 (セレンディピティ) [32], システムに対する信頼性 [34] など, 情報推薦を評価する上で重要な要素とみなされてきている. 推薦アイテムと共にその推薦理由やアイテムの特徴を提示すること (推薦説明) は, システムに対する信頼性や満足度の向上, ユーザの意思決定の支援など, 様々な効果が期待されている [35]. 図 2.8 に Amazon における推薦説明の例を示す. 図 2.8(a) では, アイテムの推薦時に推薦理由やアイテムの特徴を文章で補足説明している. 図 2.8(b) ではユーザの閲覧アイテムに基づいた推薦に対して, その推薦理由を文章と閲覧アイテムのサムネイルを用いて提示している.

Tintarev ら [35] は推薦説明が目指す推薦システムへの効果を, 表 2.1 に示した 7 つの観点からまとめている. 本節ではこの 7 つの観点を元にして, 既存の推薦説明の手法を紹介すると同時に, 有用な推薦説明, およびその評価方法について考察する.

2.3.1 Transparency (透明性)

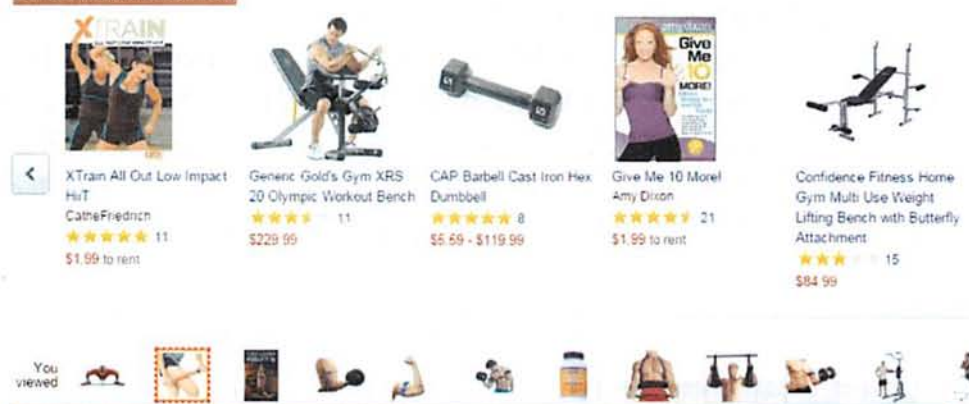
透明性とは, アイテムの推薦された過程や仕組み, 理由をユーザに理解させるものであり, 情報推薦において重要な要素であることが文献 [36] で述べられている. Cramer ら [37] は, 推薦理由を提示した後, システムの機能に対するアンケートをとることで, ユーザが本当にシステムの仕組みを理解しているかどうかを評価している.



(a) 文章による推薦説明 (下線部)

Your Recently Viewed Items and Featured Recommendations

Inspired by your browsing history



(b) 文章と画像による推薦説明 (下線部)

図 2.8: Amazon における推薦説明の提示例

2.3.2 Scrutability (理解度)

理解度とは、透明性と近い関係性にあり、アイテムが本当に正しい理由で推薦されたのかをユーザに判断できるようにする観点である [38]。Czarkowski ら [38] は休日の過ごし方の推薦時に、ユーザのプロファイルを推薦説明として提示している。例えばあるユーザに、田舎でのドライブ計画を提示した時、ユーザプロファイルとして「あなたが独身なため」、「あなたが予算を多く所持しているため」などの説明文を提示し、さらに各説明文には「why?」というリンクを載せて、その根拠を詳しく知ることができるようにしている。推薦時にユーザプロファイルやその理由と同時に提示することで、推薦に対するユーザの理解度を向上させられると報告している。

2.3.3 Trust (信頼性)

推薦されたアイテムをユーザが気に入らない場合でも、その推薦理由をユーザが理解できれば、推薦結果を許容できたり、システムに対する信頼が増す可能性がある。そのため信頼性とは透明性と関連性があり、精度や再現率とは独立した評価指標になりうるといわれている [39]。信頼性を評価する方法としては、アンケートを行ったり [40]、ロイヤリティとしてシステムに対するログイン回数を計測する方法がある [41]。

2.3.4 Persuasiveness (説得性)

推薦説明を提示することは、時として推薦アイテムを積極的に受け入れさせる効果がある。このような説得性はユーザよりもシステム側に利益をもたらすとされている [42]。Cosley ら [43] は推薦説明の提示を行うことで、ユーザの評価値をよりシステムの予測評価値に近づけることに成功している。Herlocker ら [42] は、図 2.9 に示すように対象ユーザと類似度の高いユーザが付与した評価値を、ヒストグラムや表、説明文などの形式で提示することを提案し、その有用性を示している。一方で類似ユーザが対象アイテムに下した評価を、アイテムの推薦理由として用いる手法に関しては欠点も指摘されており [44]、その詳細を 2.3.5 項にて述べる。

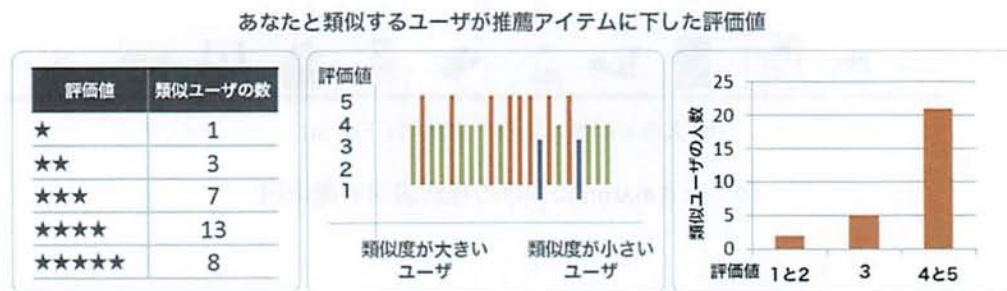


図 2.9: 類似ユーザが対象アイテムに下した評価を利用した推薦説明の提示例 (文献 [44] を元に作成)

2.3.5 Effectiveness (有効性)

説得性はユーザの推薦アイテムに対する評価を上げることが目的であるのに対し、有効性はユーザの意思決定の支援を目指す。Belgic ら [44] は、これとは対照的な手法として、説得性を重視した手法 [42] が、推薦時とアイテム使用後でユーザのアイテムに対する評価値が大きく変動することを示し、その結果が最終的にシステムの信頼性を失う恐れがあることを報告している。Belgic らは代わりに推

薦時にユーザの行動履歴をキーワードや影響度として可視化する手法を2つ提案している。1つ目はアイテムの推薦時に、そのアイテムとユーザの行動履歴との間で関連性の深い単語をリストアップする手法である。このとき各単語を選択すれば、選出された理由となった作品の一覧が提示されるようになっている。2つ目は推薦時に、そのアイテムと関連性の深い、ユーザの過去に評価したアイテムをリストアップする手法である。Belgicらはこの2つの手法と文献[42]による手法を比較し、提案手法の方がアイテム使用前後に於いて評価値が変動しないことを示している。このようにシステムに対する長期的な満足度、信頼性を考慮した場合、説得性だけでなく、ユーザの意思決定を支援する要素も重要であることがわかる。

有効性を評価する方法としては、前述のようにアイテムの使用前後の評価値を比較する方法が用いられる。異なる評価方法として、Vigら[45]は表示された推薦説明に対して「この説明文はこの映画を評価する上で判断に役立ったか？(This explanation helps me determine how well I will like this movie)」といった質問を5段階で評価してもらう手法を提案している。

Cramerら[42]は有効性を精度にとって変わる評価指標の1つとして提案しており、Thompsonら[46]は有効性が精度、再現率にもっとも近い関係がある指標であると述べている。このことから情報推薦においては、精度だけでなく有効性の観点からシステムを評価することが重要であることがわかる。

2.3.6 Efficiency (効率性)

推薦説明の効果の1つに、ユーザの意思決定を早めることが挙げられる。McCarthyら[47]はアイテム閲覧時に、「さらに～の機能をもった商品はこちらになります」と関連アイテムを表示することで、ユーザの検索効率を上げるを試みている。効率性の評価手段としては、目的アイテムを探し出すまでの時間の評価[46]、目的のアイテムを探せなかった場合にユーザが要求したフィードバックの分析、ユーザが閲覧した推薦説明の数の集計などがある[?,48]。

2.3.7 Satisfaction (満足度)

推薦説明を提示することで、ユーザがシステムに満足する効果や、システム利用における飽きや苦痛を低減させる効果が期待できる[48]。満足度の評価方法には、システムに対して満足したかどうかをアンケートで尋ねたり、提示された推薦説明に対して評価を下す方法がある[49,50]。また実験の様子を観察する方法も存在し[51]、コストはかかるものの具体的なシステムに対するユーザの反応を詳しく分析することが可能と考えられる。2.3.3項で述べたように、推薦精度とユーザのシステム自体に対する評価は必ずしも一致するわけではなく、推薦アイテム自体の評価とシステムに対する評価を分けて行うことが重要とされている[37,48]。

3 価値観アイテムモデリング手法

3.1 アイテムモデリングの利用

本論文で提案する価値観アイテムモデリング手法は、2.2節で紹介した価値観に基づいたユーザモデリング手法 [2] を拡張したものであり、ユーザではなくアイテムをモデリングする。

価値観ユーザモデリングでは、価値観を「ユーザの属性に対するこだわり」とした。それに対し本論文では、価値観を「アイテムにおける評価観点」と考え、アイテムのモデルを作成することを提案する。図 3.1 にアイテムにおける評価観点の例を示す。ここではテレビに対する意見として、3 人のレビューが付与された場合を

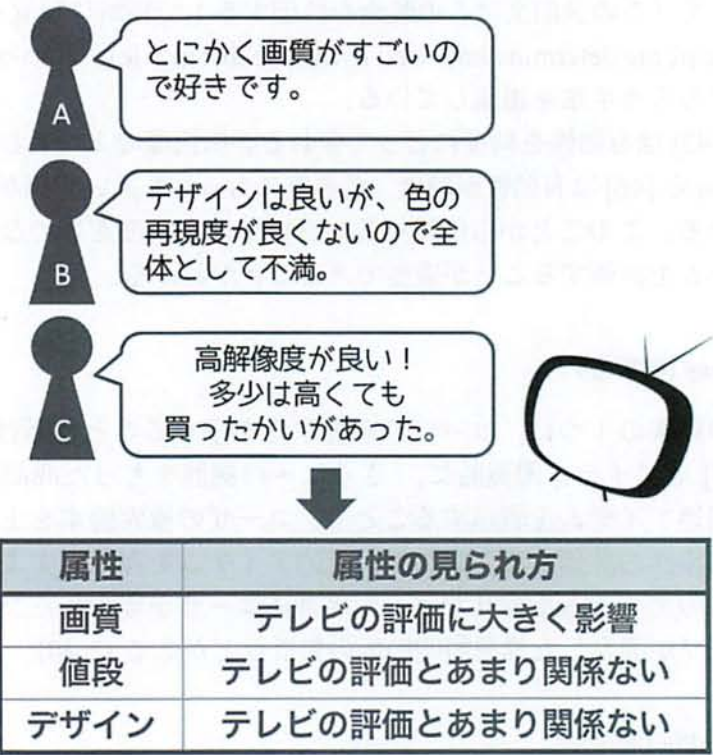


図 3.1: アイテムにおける評価観点の例

想定している。この場合、レビューよりユーザがデザインや値段に不満をもっているとしても、画質を気に入った場合にテレビを好評としている傾向があることがわかる。このことから、値段やデザインよりも画質の評価がテレビの評価に大きく影響していることが推論される。提案手法では、このようにユーザ個人が投稿したレビューを収集・分析することで、アイテムがどのように見られているのか、そ

の傾向をアイテムの総合評価と属性に対する評価との関係から推論する。

このようなアイテムの評価観点を推論できれば、例えばある属性にこだわりを持つユーザがいた場合、その属性に関する特徴 (総合評価への影響の大きさなど) を一緒に提示することで、多くの人から低評価を受けていても対象ユーザが気に入る可能性のあるアイテムを推薦することが可能になり、多様性やセレンディピティの向上に繋がるのが期待できる [6]。

また似たような評価観点を持つアイテムを分類し、ユーザに提示することもできる。さらに高評価のアイテムでも、提示する評価観点によっては、反対に消極的な情報を付加することも可能である。推薦時にアイテムに関する消極的な情報をユーザに提示することは一般的には問題とされるが、2.3 節で述べたように適切な見せ方を選択することで、推薦の信頼性の向上に繋がる可能性がある。

図 3.2 に価値観に基づいたユーザモデリングと、提案するアイテムモデリングの比較を示す。価値観ユーザモデリングでは、あるユーザ u_i が評価した全アイテムのレビューを元に RMRate を用いてユーザモデルを構築する。一方アイテムモデリングでは、あるアイテム i_t に評価されている全ユーザのレビューを元にアイテムモデルを構築する。このことから価値観に基づいてアイテムモデルを作成することは、ユーザモデリングと比較して以下の利点が期待できる。

- アイテムあたりのレビューを多数得やすいため、より詳細なモデリング (リフト値の利用) が可能
- アイテムモデルをアイテム間の類似度計算に使用することで、アイテムベースの推薦システムに適用可能
- アイテムモデルをアイテムの特徴として扱うことで、推薦の説明が可能

本章ではこれらの利点に基づいて、3.2 節では RMRate の拡張 (リフト値の利用) について、3.3 節ではアイテムベースの推薦システムへの適用方法、3.4 節では推薦説明への適用方法についてそれぞれ詳細を述べる。

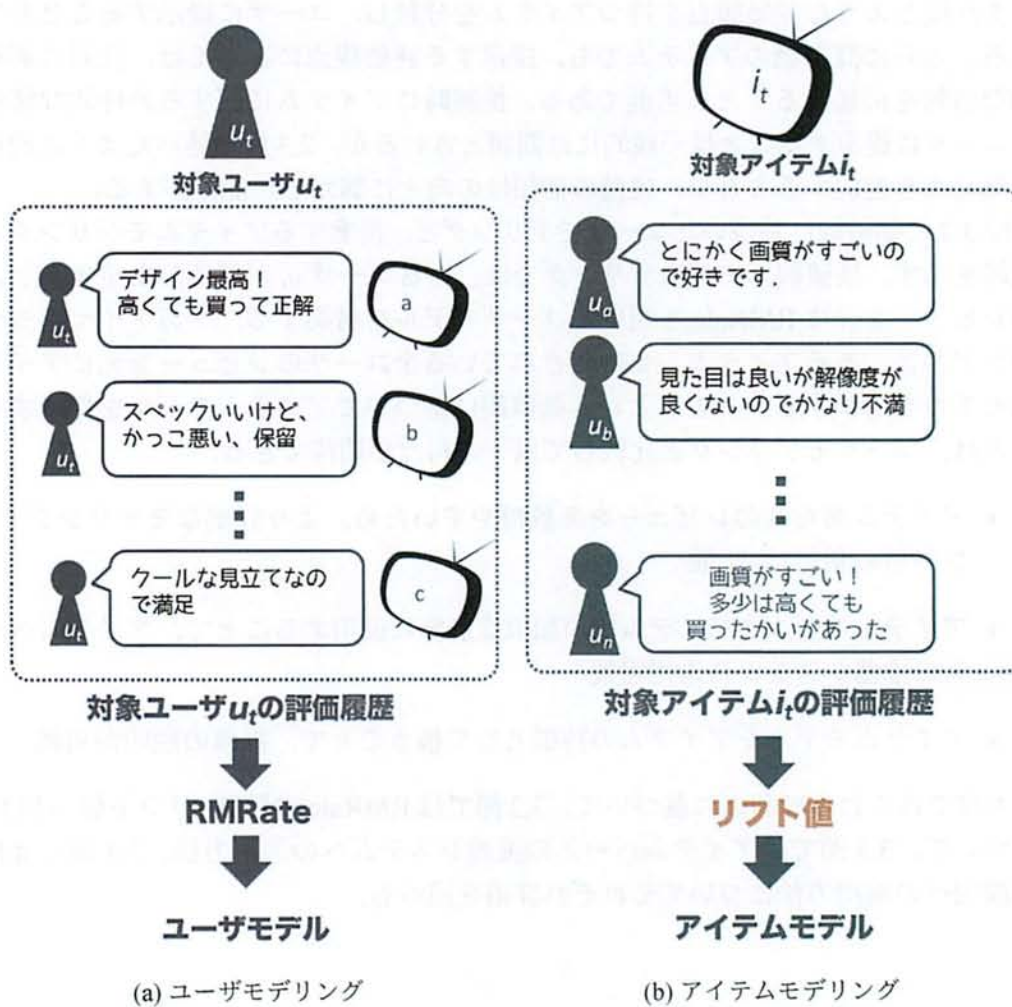


図 3.2: 価値観ユーザモデリングと提案するアイテムモデリングの比較

3.2 RMRate の拡張

2.2 節で述べたとおり，RMRate ではアイテムに対する評価と，属性に対する評価の極性が一致した回数に着目している．すなわち，好評で一致した場合と不評で一致した場合を区別せずに扱っている．これにより，少数のレビューからでもモデル構築が可能である．しかし 3.1 節で述べたようにアイテムモデリングではユーザモデルと比べ，モデル構築に多数のレビューを利用可能という利点がある．そこで提案手法では，アイテムに対する好評・不評，属性に対する好評・不評の組み合わせ 4 種類に分類してモデル化を行う．具体的にはアイテムとその属性の評価極性について，極性が一致したかではなく，両者の関係を相関ルールとして検出することで，アイテムの特性をより詳しく分析する．

3.2.1 相関ルールについて

相関ルールとはデータマイニングの 4 つであり，アイテム間における組み合わせの規則を解析する手段である [52]．元々はスーパーマーケットでの買い物かごの内容を分析することで，マーケティングに役立てようとするバスケット分析を目的に提唱された方法論であるが，その仕組みが他のデータ分析に適用できる汎用性の高いものであるため，今日でも様々な分野に利用されている [52]．相関ルールは，大規模なデータを分析することで，関係性の強い事象の組み合わせや，そのような関係を表す規則を見つけ出す．相関ルールでは，あるアイテムの組み合わせで構成される 1 つの集合をトランザクションと呼ぶ．マーケットにおける一人の顧客が購買したアイテムのリストなどがトランザクションの例に挙げられる．相関ルールはアイテム集合 X を条件部，アイテム集合 Y を結論部として「 $X \Rightarrow Y$ 」と表記され，代表的な指標としては支持度 (support)，確信度 (confidence)，リフト値 (lift) などがある．

支持度の計算式を式 3.1 に，確信度の計算式を式 3.2 にそれぞれ示す．

$$\text{supp}(X \Rightarrow Y) = \frac{\text{freq}(X, Y)}{N} \quad (3.1)$$

$$\text{conf}(X \Rightarrow Y) = \frac{\text{freq}(X, Y)}{\text{freq}(X)} \quad (3.2)$$

ここで $\text{freq}(X)$ は X を含むトランザクションの件数， $\text{freq}(X, Y)$ は X と Y を含むトランザクションの件数， N は全てのトランザクション件数を意味する．支持度はルール「 $X \Rightarrow Y$ 」の出現確率，確信度は X と Y の結びつきの強さを表し，それぞれ 0 から 1 までの値を持つ．

図 3.3 に支持度，確信度の計算例を示す．映画を鑑賞した人のうち，図 3.3(a) では映画 A と映画 B を好評とした人，図 3.3(b) では映画 A と映画 C を好評とした人についての，それぞれの支持度，確信度を計算している．図 3.3(b) のように結論

部 Y を含むトランザクションが大きい場合、 X との関係とは独立に $freq(X, Y)$ が大きくなりやすいため、支持度、確信度が大きくなる傾向がある。つまり結論部のアイテムが人気アイテムなどで単独の出現確率が高い場合、条件部によらず支持度や確信度は高く評価される傾向になる。

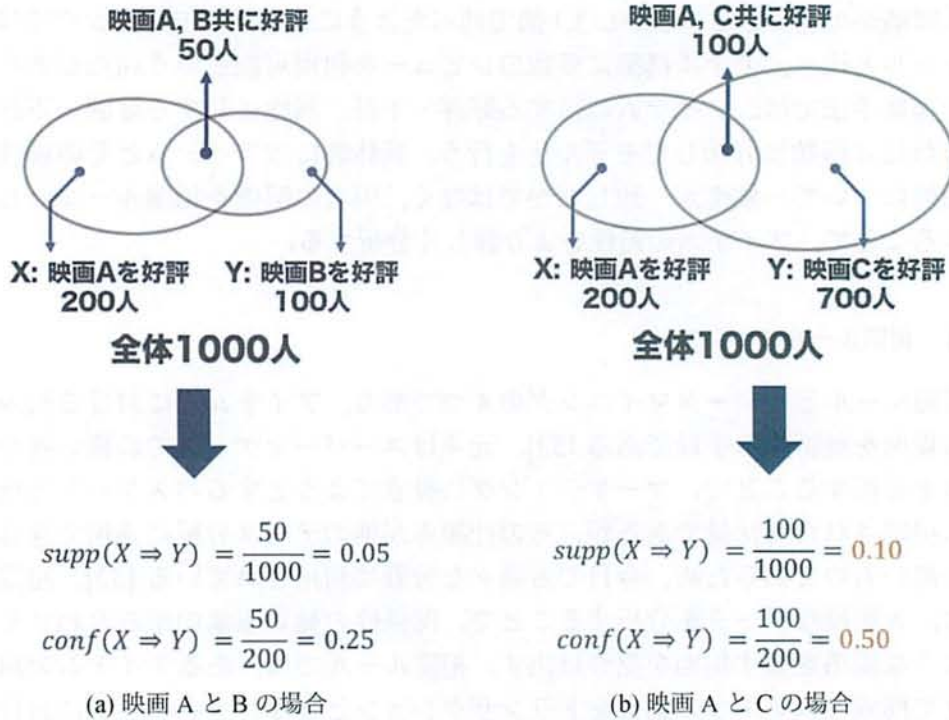


図 3.3: 映画の評価における支持度、確信度の計算例

3.2.2 リフト値

図 3.3 の例で示した確信度の欠点を補う方法の 1 つにリフト値があり、以下の式 3.3 により計算される。

$$lift(X \Rightarrow Y) = \frac{conf(X \Rightarrow Y)}{P(Y)} = \frac{P(X \cap Y)}{P(X)P(Y)} \quad (3.3)$$

ここで $P(X)$ は X の出現確率を意味する。式 3.3 よりリフト値は、確信度を結論部の出現確率で割ったものであり、 Y の出現確率が条件 X により何倍増加するかを意味する。図 3.4 に、図 3.3 と同じ条件下でのリフト値の計算例を示す。図 3.4(a) におけるリフト値は「映画 B を好評とする確率は、映画 A を好評とすることにより 2.5 倍増加する」ことを意味している。また図 3.4 より、図 3.4(a) の方が図 3.4(b) よりリフト値が大きくなることがわかる。このようにリフト値を使用することで、少数でも結びつきの強い組み合わせを検出することが可能となる。

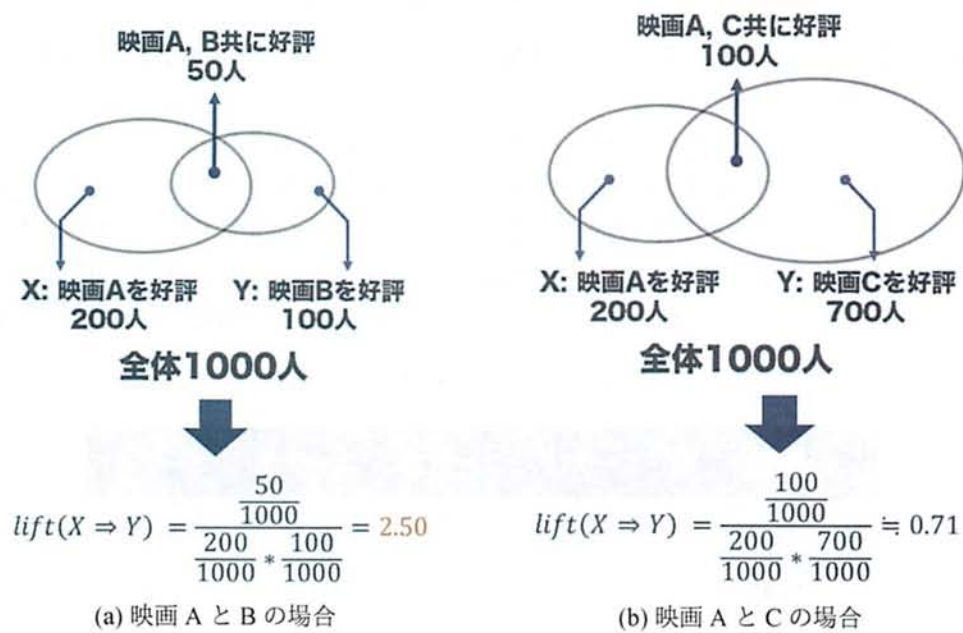


図 3.4: 映画の評価におけるリフト値の計算例

3.2.3 リフト値によるモデリング方法

提案手法では、リフト値を使用してアイテムモデルを求める。相関ルールでは、 X, Y はあるアイテム集合がトランザクションに含まれる事象を表していた。提案手法では属性に対する評価、アイテムに対する評価が X, Y にそれぞれ対応する。好評 (positive), 不評 (negative) をそれぞれ「pos」, 「neg」と表した場合、アイテム i に対するレビューの評価極性が $p_i(\in \{pos, neg\})$ となる事象 X_i^r と、アイテム i に対するレビューにおいて属性 j の評価極性が $p_a(\in \{pos, neg\})$ となる事象 Y_{ij}^a の間のリフト値 $lift_{ij}(p_a \Rightarrow p_i)$ は式3.4で定義される。

$$lift_{ij}(p_a \Rightarrow p_i) = \frac{P(Y_{ij}^a \cap X_i^r)}{P(Y_{ij}^a)P(X_i^r)} \quad (3.4)$$

例えば $lift_{ij}(pos \Rightarrow pos)$ の値は、 i が好評となる確率が、属性 j が好評の場合にどのくらい高くなるかを意味している。

式3.4を用いて、 p_a, p_i の組み合わせにより属性毎に4通りのリフト値が求められる。例として、ある映画に対して4人のユーザがある映画の評価を行った例を表3.1に示す。また、表3.1の評価に基づいて属性ごとに4通りのリフト値を計算した例を表3.2に示す。例えば1列目($pos \Rightarrow pos$)は、属性に対する評価が好評の時に総合評価も好評となる場合のリフト値を意味する。この例では、属性「物語」

表 3.1: アイテムにおける評価例

属性	ユーザ1	ユーザ2	ユーザ3	ユーザ4
総合	neg	neg	pos	neg
物語	neg	neg	pos	pos
映像	pos	pos	neg	neg

表 3.2: 表 3.1 に対するリフト値の計算結果

属性	$pos \Rightarrow pos$	$pos \Rightarrow neg$	$neg \Rightarrow pos$	$neg \Rightarrow neg$
物語	2.00	0.67	0	1.33
映像	0	1.33	2.00	1.33

が好評だとその映画も好評になる傾向にあるといえる。また、属性「映像」に関しては不評の場合にその映画は好評になるといえる。このようにリフト値を計算することで、総合評価に影響を与える属性を求めるだけでなく、消極的な情報を

含めた，総合評価とその属性の間に存在するより詳細な関係性を推論することが可能となる．

3.3 アイテムモデルのアイテムベースによる推薦への適用

本論文では3.2節で作成したアイテムモデルをアイテムベースの推薦へ適用することを提案する。ここでは、アイテムモデルを Sarwar ら [15] のアイテムベース推薦手法へ適用する手順を説明する。

3.3.1 リフト値の類似度計算への適用方法

2.1節で述べたように、アイテムベースの推薦では始めにアイテム間の類似度を算出する。文献 [15, 16] では、類似度の計算に対象アイテムに下された評価値を利用している。提案手法では、代わりに3.2節で計算される各属性毎の4種類のリフト値を利用する。基本的には、アイテムから計算される全てのリフト値をまとめて1つのベクトル (属性の数を m とした場合、 $4m$ 次元のベクトル) として扱い、そのベクトル同士の類似度を求める。

上記の他、各属性毎、または p_a, p_t の組み合わせ毎にリフト値をまとめ、類似度を計算する方法も考えられる。例えば p_a, p_t の組み合わせによる4種類のリフト値を、各属性毎に4次元のベクトルとしてまとめ、それぞれの類似度を計算した場合、類似しているアイテム同士が、さらにどの属性において類似しているかを推論できる。例えばある2つの映画において、物語におけるリフト値の類似度が大きい場合、両映画は物語における評価のされ方で類似、または物語の総合評価への影響の仕方で類似していることがわかる。ただしこの手法では、比較対象であるベクトルの次元数が少ないため、Pearson 相関などでは適確な類似度の計算が困難になることが想定される。

また全 m 個の属性から計算されるリフト値を、 p_a, p_t の組み合わせ別に m 次元のベクトルとしてまとめ、それぞれ類似度を求める方法の場合、似ているアイテム同士の、さらに属性に対する共通の見られ方を推論することができる。例えばある2つの映画において、リフト値 ($pos \Rightarrow pos$) での類似度が大きい場合、両映画とも、その映画の持つ属性を好評とした人は、総合評価も好評にしているという観点で類似していることがわかる。しかし、この方法もアイテムの持つ属性が少ない場合、適確な類似度計算は難しくなることが想定される。

このように類似度の計算にリフト値をどのように使用するかは、目的やデータセットの状況に応じて考察する必要があると考える。4章では、ここで述べた複数の類似度計算方法について予備実験を行い、その結果に基づき考察する。

3.3.2 アイテム間類似度の計算方法

3.3.1 項で述べたアイテムモデルのベクトルから類似度を計算する方法を説明する。ここでは、アイテムから算出される全リフト値の集合を1つのベクトルとし、類似度の計算を行う。アイテムベースでは、コサイン距離より Pearson 相関を利用して推薦を行った方が MAE が少なくなるとの報告 [15] があるため、ここでは Pearson 相関を採用する。アイテムモデルを利用した Pearson 相関によるアイテム i_α と i_β の類似度計算式を式 3.5 に示す。

$$sim_{item}(i_\alpha, i_\beta) = \frac{\sum_{L_{\alpha\beta}} (l_{\alpha jk} - \bar{l}_\alpha)(l_{\beta jk} - \bar{l}_\beta)}{\sqrt{\sum_{L_{\alpha\beta}} (l_{\alpha jk} - \bar{l}_\alpha)^2} \sqrt{\sum_{L_{\alpha\beta}} (l_{\beta jk} - \bar{l}_\beta)^2}} \quad (3.5)$$

ここで k は p_a, p_t の組み合わせによるリフト値の種類を意味し、 $L_{\alpha\beta}$ は i_α と i_β に共通して存在するリフト値の集合を表す。 $l_{\alpha jk}$ は i_α での属性 j における種類 k ($pos \Rightarrow pos, pos \Rightarrow neg, neg \Rightarrow pos, neg \Rightarrow neg$ のいずれかに対応) のリフト値を示し、 \bar{l}_α は i_α で計算された全リフト値の平均値を表す。図 3.5 に類似度の計算例を示す。計算結果より映画 A と映画 B の相関は中くらいであるとわかる。

求めた類似度をアイテムベースの協調フィルタリングに応用する場合には、式 3.5 で求めた類似度を、式 2.4 に適用することでアイテムの予測評価値を求めることが可能となる。

映画Aのリフト値の計算結果

属性	$pos \Rightarrow pos$	$pos \Rightarrow neg$	$neg \Rightarrow pos$	$neg \Rightarrow neg$
物語	2.00	0.67	-	1.33
映像	0	1.33	2.00	1.33

映画Bのリフト値の計算結果

属性	$pos \Rightarrow pos$	$pos \Rightarrow neg$	$neg \Rightarrow pos$	$neg \Rightarrow neg$
物語	1.14	0.33	0	2.14
映像	0.33	-	1.00	-

$$\begin{aligned}
 &sim_{item}(\text{映画A}, \text{映画B}) \\
 &= sim_{item}((2, 0.67, 1.33, 0, 2), (1.44, 0.33, 2.14, 0.33, 1)) \\
 &\doteq 0.5488
 \end{aligned}$$

図 3.5: アイテムモデルの類似度の計算例

3.4 アイテムモデルの推薦説明への適用

本節では、作成したアイテムモデルを推薦説明へ利用することを提案する。価値観に基づいたアイテムモデルは、そのアイテムの評価観点を推定した結果を表している。そこで推薦時やアイテムの閲覧時に、アイテムモデルから作成した説明文をそのアイテムの特徴として提示することにより、アイテムの魅力を評価値以外の形で提供することが可能と考える。これにより、ユーザの意思決定を支援する効果が期待できる。

アイテムモデルを元にした説明文の作成、提示方法はいくつかの方法が考えられる。例えば図3.6のように、各属性のリフト値を直接提示する方法がある。図3.6ではYahoo!映画に存在するある映画から構築したアイテムモデルを、表とレーダーチャートにて可視化している。この様な提示方法では、アイテムの詳細な情報を客観的に伝えることが可能である。一方でユーザはリフト値について理解する必要があり、また提示された情報量が多いため、ユーザが提示内容を理解するための負担は大きい。

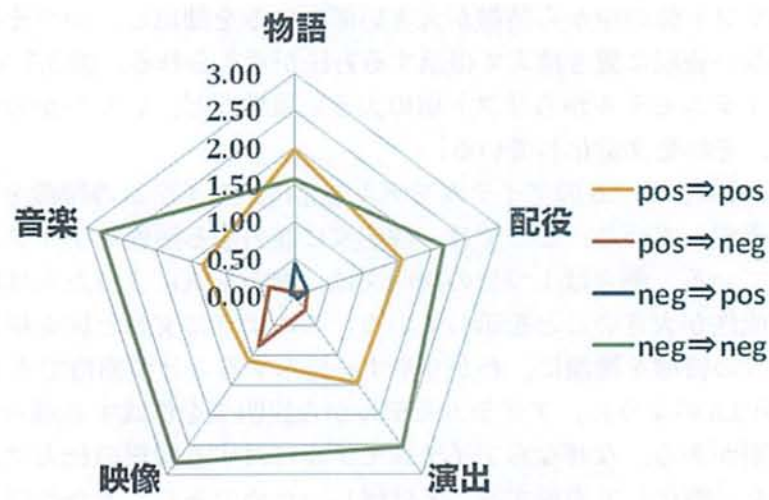
そこで各リフト値の中から特徴が大きい部分のみを抽出し、かつそれらを専門知識を使わない表現に置き換えて提示する方法が考えられる。図3.7では、図3.6で示したアイテムモデルからリフト値の大きい属性など、いくつかの目立つ特徴を取り上げ、それを文章化している。

図3.8も同様に、図3.6のアイテムモデルを元に、アイテムの特徴を文章と図を組み合わせて表現している。ここでは、説明文に該当する属性をアイコンのような形で表示している。例えば1つ目の説明文は、物語を気に入った人は映画自体を気に入る可能性が大きいことを示している。このように文章と図を併用することで、アイテムの特徴を簡潔に、わかりやすく提示することが期待できる。

図3.7や図3.8のように、アイテムモデルから説明文を作成する場合、十分な考察を行う必要がある。なぜならアイテムモデルに対する解釈の仕方は人により異なり、それを一般化して表現することは難しいためである。また作成した説明文がユーザにとって理解しやすいものである必要もある。アイテムモデルに対する解釈やその表現方法については4章で実験結果に基づき考察する。

属性	$pos \Rightarrow pos$	$pos \Rightarrow neg$	$neg \Rightarrow pos$	$neg \Rightarrow neg$
物語	1.96	0.04	0.43	1.57
配役	1.57	0.18	0.18	2.18
演出	1.46	0.24	0.06	2.53
映像	1.09	0.84	0.00	2.80
音楽	1.33	0.37	0.06	2.79

(a) 表による提示



(b) レーダーチャートによる提示

図 3.6: アイテムモデルによる直接的な推薦説明の提示例



この映画の特徴

この映画の物語を気に入って頂けた場合、映画自体を気に入って頂けるでしょう。一方で配役、演出、音楽、映像に不満を感じた人は、作品自体を不満に感じる傾向にあります。

図 3.7: アイテムモデルを元にした，文章による推薦説明の提示例



この映画の特徴

ここを気に入った人は映画も好きになれるでしょう。

物語

ここに不満を感じた人は、映画を不評としている傾向にあります。

配役

演出

音楽

映画

図 3.8: アイテムモデルを元にした，文章と図による推薦説明の提示例

4 評価実験

本章では、提案手法の有効性を検証するため、プロトタイプシステムを構築して評価実験を行う。提案システムでは、3章で提案したアイテムモデルを、推薦アルゴリズムと、推薦説明の両方に利用する。アイテムモデルの解釈および推薦説明文について検討するために、3.4節で述べたように、予備実験を事前に行った結果を4.1節に示し、評価実験については4.2節にて述べる。

4.1 アイテムモデルの解釈に関する予備実験

4.1.1 実験概要

本節では、提案するアイテムモデリングに基づく推薦説明について、その生成と提示方法の考察のために行った予備実験を説明する。実験では、工学系大学生および大学院生14名に対し、提案するアイテムモデリングの結果を提示し、結果に対する解釈やこれを利用したアイテムの説明文等をアンケート形式で収集した。

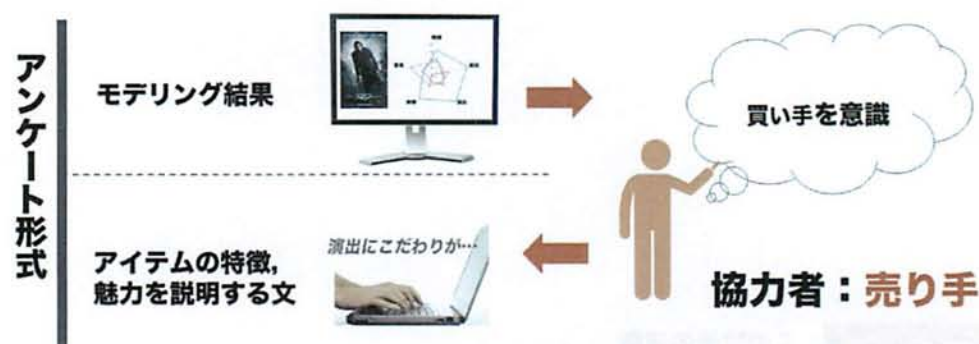


図 4.1: 予備実験の概要図

アンケートでは図 4.1 に示すように、属性値やサムネイル画像といったアイテムの基本情報とモデリング結果を提示し、実験協力者に「売り手として、アイテムの魅力を買い手に宣伝する」ことを想定し、提示された情報を元にそのアイテムの魅力や特徴を説明する文を記述してもらった。

アイテムモデリングの結果としては、各属性における $(Y_{ij}^a \cap X_i^r)$ (3.2.3 項参照) を満たすレビューの数と、4 種類のリフト値の値を表およびレーダチャートにしたものを提示した。モデリングには、旅行サイト 4travel¹ より宿泊施設 3 件（アイテム 1, 2, 3）、Yahoo!映画より映画 2 件（アイテム 4, 5）の合計 5 件のアイテムを利

¹<http://4travel.jp/>

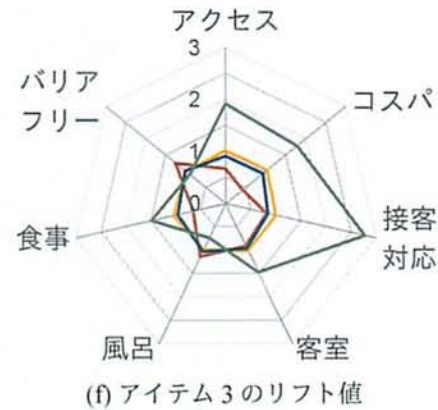
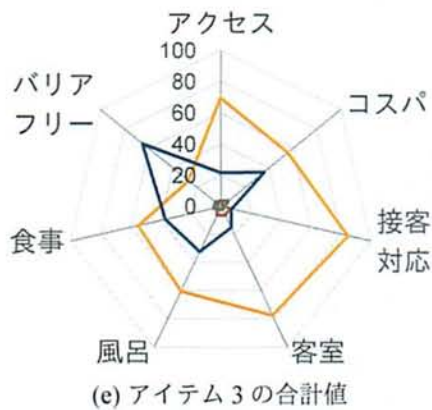
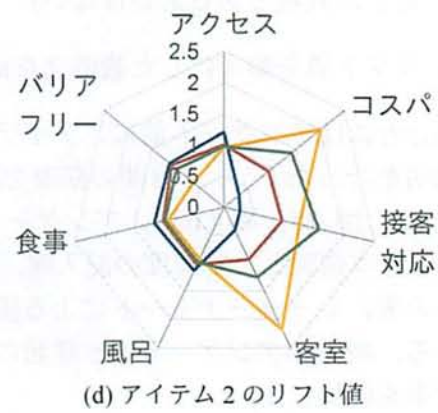
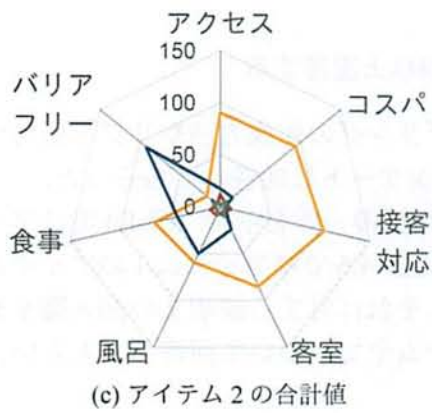
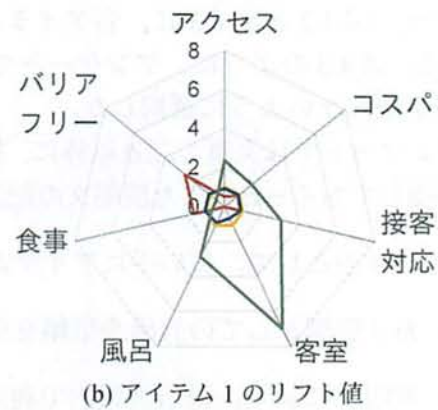
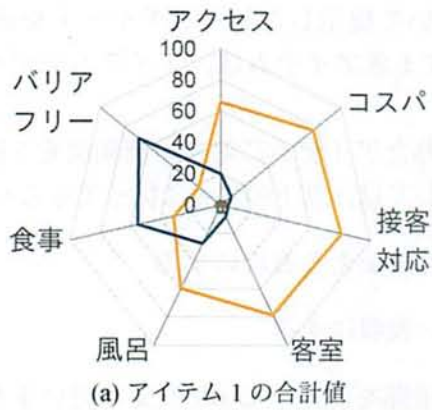
用した。図 4.2 と図 4.3 に、各アイテムについて提示したレーダチャートを示す。図 4.2, 図 4.3 のように、アンケートで提示する各アイテムは、アイテムモデルの特徴が被らないように選択した。

アンケートでは文章の記述以外に、提示されたアイテムに対する認知度を 5 段階で評価してもらった。また説明文の記述に関しては、以下の 4 点に従ってもらった。

- 売り手として、買い手にアイテムの魅力を伝える表現にする
- 売り手側としての評価や信頼を失わない表現にする
- 説明文に加え、提示情報内で利用した根拠を記述する (理由文は買い手を想定した表現である必要はない)
- リフト値を参考にした説明文を必ず 3 つ以上記述する

協力者にはアンケート前に、アイテムモデリングの作成方法やリフト値について説明を行った上で、2 週間の期限を設けアンケートに回答してもらった。

図 4.4, 図 4.5, 図 4.6 に、アンケート回答用紙の一部を示す。図 4.4 ではアイテム 1 の基本情報と、認知度の記入欄、図 4.5 と図 4.6 ではアイテム 1 のアイテムモデルの表、レーダーチャートによる提示と、それに対する説明文の記入欄を載せている。同様のアンケートを 5 種類のアイテム全てについて回答してもらい、その結果を収集した。



— 属性評価：好評 総合評価：好評
 — 属性評価：好評 総合評価：不評
 — 属性評価：不評 総合評価：好評
 — 属性評価：不評 総合評価：不評

(g) 各グラフの凡例

図 4.2: 実験にて提示したアイテム情報(レーダーチャート) その1

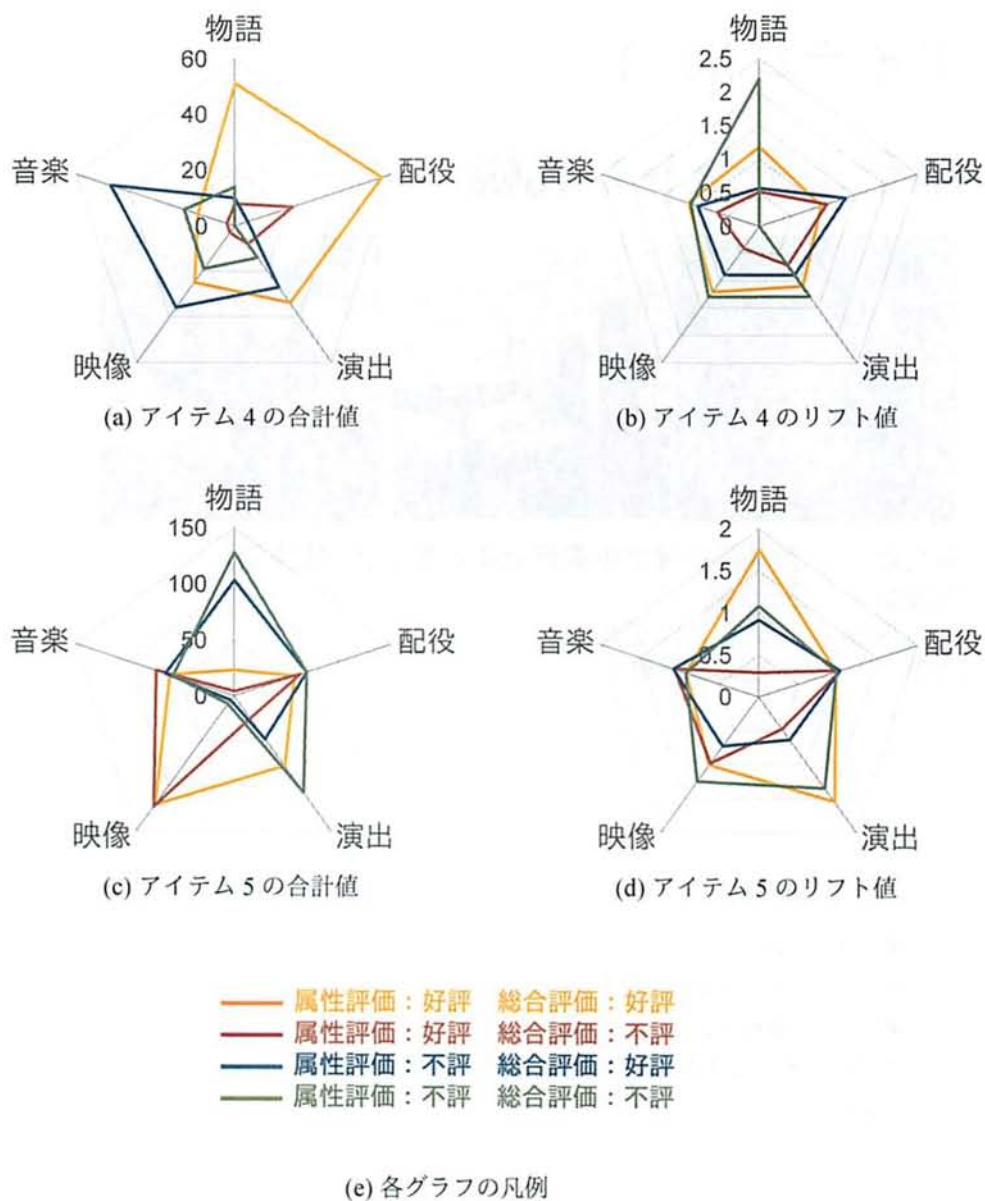


図 4.3: 実験にて提示したアイテム情報(レーダチャート) その 2

アンケート回答用紙

氏名

アイテム No. 1

ホテルモントレ札幌



所在地 : 北海道札幌市中央区北4条東1-3 (札幌駅)
客室数 : 250室
施設形態 : ビジネスホテル
開業 : 1994年6月

このホテルをご存知ですか？

1. 全く知らない
2. 名前を聞いたことがある程度
3. 実際に、1回利用したことがある
4. 何回も(2回以上)利用したことがある
5. その他

1～5の番号を記入

その他(5)を選択した人は説明をお願いします

図 4.4: アンケート回答用紙の一部その1: アイテムの基本情報の提示と、認知度の記入欄

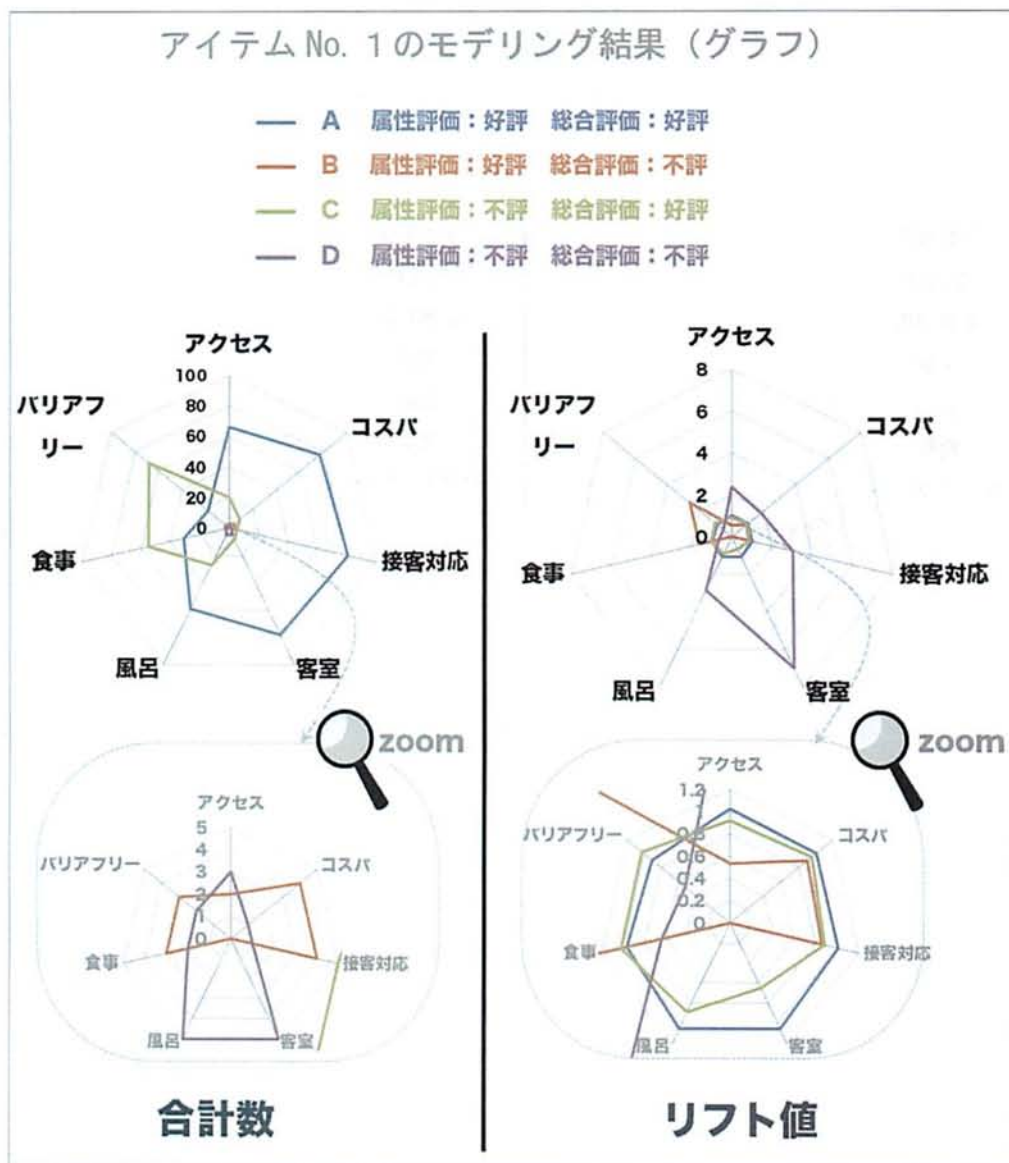


図 4.5: アンケート回答用紙の一部その2：レーダーチャートによるアイテムモデルの提示

アイテム No. 1 のモデリング結果 (表)				
— A 属性評価：好評 総合評価：好評				
— B 属性評価：好評 総合評価：不評				
— C 属性評価：不評 総合評価：好評				
— D 属性評価：不評 総合評価：不評				
レビュー数：91 件				
	A	B	C	D
アクセス	66	2	20	3
コスバ	77	4	9	1
接客対応	81	4	5	1
客室	78	0	8	5
風呂	59	0	27	5
食事	31	3	55	2
バリアフリー	18	3	68	2
合計数				
アイテム No. 1 の説明文				
	A	B	C	D
アクセス	1.03	0.54	0.92	2.37
コスバ	1.01	0.90	0.95	1.82
接客対応	1.01	0.86	0.88	3.03
客室	1.06	0.00	0.65	7.00
風呂	1.06	0.00	0.89	2.84
食事	0.96	1.61	1.02	0.64
バリアフリー	0.91	2.60	1.03	0.52
リフト値				

図 4.6: アンケート回答用紙の一部その 3：表によるアイテムモデルの提示と、説明文の記入欄

4.1.2 実験結果

アンケート結果より、アイテムに対する認知度をまとめたグラフを図 4.7 に示す。図 4.7 より、対象アイテムに対する利用経験は少なく、従って協力者はほぼア

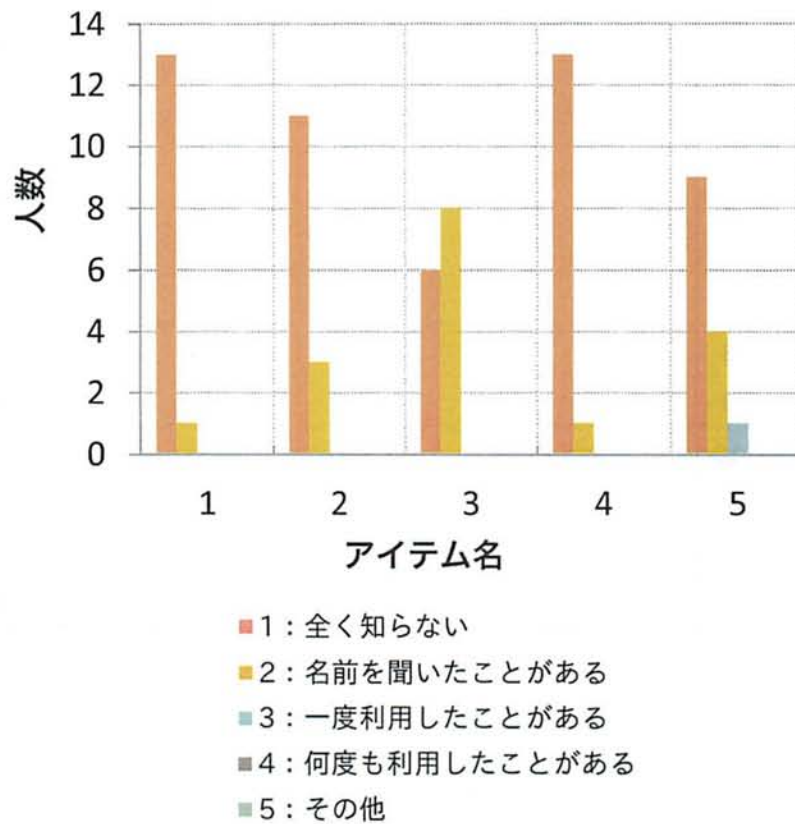


図 4.7: アイテムに対する認知度のグラフ

ンケートで提示された情報からアイテムの特徴を捉え、説明文を作成していると考えられる。アイテムモデリングに対する説明文は、協力者 14 名より 221 個の文章が得られた。図 4.8 に実際に得られた回答の一部を示す。

得られた文章について、「リフト値が大きいこと」を根拠として作成された説明文は 190 個存在した。その中には図 4.8(b) や、図 4.8(c) の 2 つ目の文章のように、「『属性』がおすすめ、優れている、評価されている」等の共通した説明文が見られた。そこで回答された説明文を、類似している内容に基づいて以下の 13 種類に分類した。

1. 「属性」が好評(おすすめ)である。
2. 「属性」を好評とした人は、アイテムを気に入る(アイテムをおすすめできる)。
3. 「属性」にこだわり(興味)を持つ人は、アイテムを気に入る。

表 4.1: リフト値が大きいことを根拠とした説明文の分類結果

説明文を 記述した理由		$lift_{ij}(pos \rightarrow pos)$ が大きい	$lift_{ij}(pos \rightarrow neg)$ が大きい	$lift_{ij}(neg \rightarrow pos)$ が大きい	$lift_{ij}(neg \rightarrow neg)$ が大きい
文章の種類 (Index)	1	35	3	2	6
	2	7	1	1	0
	3	19	2	0	1
	4	0	0	1	18
	5	0	1	0	1
	6	2	4	3	21
	7	1	0	0	2
	8	0	1	0	1
	9	0	0	0	0
	10	2	11	15	2
	11	0	0	0	0
	12	0	7	5	0
	13	9	2	2	8

4. 「属性」が不評である。
5. 「属性」を不評とした人は、アイテムを気に入らない(アイテムをおすすめできない)。
6. 「属性」にこだわりを持つ人は、アイテムを気に入らない。
7. 「属性」がアイテムの評価に影響を与えている。
8. 「属性」を好評とした人は、アイテムを気に入らない。
9. 「属性」を不評とした人は、アイテムを気に入る。
10. 「属性」にこだわりを持たない人でも、アイテムを気に入る。
11. 「属性」にこだわりを持たない人は、アイテムを気に入らない。
12. 「属性」がアイテムの評価に影響を与えていない。
13. その他

また回答された説明文の分類と、その時根拠にされたリフト値の種類について、その頻度を集計した結果を表 4.1 に示す。ここで 1 行目の文章の種類 (Index) は、分類した 13 種類の文章の番号に対応し、例えば 1 行 1 列目は「 $lift_{ij}(pos \Rightarrow pos)$ (3.2.3 項参照) の値が大きい」という理由で、「『属性』はおすすめである」と表現した説明文が 35 件あったことを示している。各リフト値を根拠とした説明文の中で、頻度上位 2 件の文章の種類を赤字で示している。アンケート回答例を図 4.8 に示す。

表 4.1 より, $lift_{ij}(pos \Rightarrow pos)$ の値が大きいという理由から (1) 「『属性』が好評である」, (3) 「『属性』に対してこだわりを持つ人は, アイテムを気に入る」という表現が多く作成されたことがわかる. また $lift_{ij}(neg \Rightarrow neg)$ の値が大きいとき, (4) 「『属性』が不評である」, (6) 「『属性』にこだわりを持つ人は, アイテムを気に入らない」という解釈のされ方が多いことがわかる. 属性評価と総合評価の極性が同じ場合のリフト値が大きいとき, (1), (4) の文章が多く見られたのは, リフト値に加えて $(Y_{ij}^a \cap X_i^a)$ の数で判断された可能性が考えられる. アンケートに対する回答でも, 図 4.8(a) の 1 件目のようにリフト値だけでなく評価件数も参考にした回答も複数見られた. また $lift_{ij}(pos \Rightarrow pos)$ の値が大きいとき, (3) の文章も多く見られたのは, リフト値の持つ意味に基づいていると考えられる. つまり $lift_{ij}(pos \Rightarrow pos)$ が大きいということは, その属性を気に入った場合, 総合評価を好評とする確率が大きくなることを意味するため, 「こだわり, 興味を持つ人」などの条件を用いた表現が使用されたと考えられる. $lift_{ij}(neg \Rightarrow neg)$ の値が大きいとき, (6) の文章が多く見られた理由についても同様のことが言えると考える.

複数のリフト値を考慮した説明文も存在した. 例えば図 4.8(c) の 1 件目の説明文のように, $lift_{ij}(pos \Rightarrow pos)$ と $lift_{ij}(neg \Rightarrow neg)$ の値が両方とも大きい時, 「このアイテムは『属性』によって評価が分かれる, 好き嫌いの大きいアイテムである」等の説明文が多く見られた. この場合, (1) と (4) の状態が同時に存在するため「好き嫌いが大きい」のような表現が使われたり, (3) と (6) の両方の内容を表現するため「『属性』によってアイテムの評価が分かれる」などの言い方がされたのだと考える. 表 4.1 において, これらの文章は (13) その他に分類した. $lift_{ij}(pos \Rightarrow pos)$ と $lift_{ij}(neg \Rightarrow neg)$ において「その他」に分類されている説明文が多いのはそのためである.

$lift_{ij}(pos \Rightarrow neg)$ や $lift_{ij}(neg \Rightarrow pos)$ の値が大きいという理由により, (10) 「『属性』にこだわりを持たない人でも, アイテムを気に入る」や (12) 「『属性』がアイテムの評価に影響を与えていない」という表現が多くされている. これらは属性の評価がアイテムの評価に与えている影響は小さいという解釈に基づくものと言える. また, $lift_{ij}(pos \Rightarrow neg)$ については, それに基づく説明文生成が難しいと言う意見が多く協力者から寄せられた. そのため $lift_{ij}(pos \Rightarrow neg)$ に対するリフト値の解釈は難しいと考えられる.

回答された説明文の文中には, 「リフト値」や「確率」といった用語を使用した表現は一切見られなかった. この結果は, 協力者が売り手側としてアイテムの宣伝を考慮し, 直接的な表現を控えたためと考える.

回答された説明文の中には, 推薦システムで利用可能な有用な表現も存在した. 例えば図 4.8(a) の 3 件目の説明文のように, $lift_{ij}(neg \Rightarrow neg)$ の値が大きいとき, 「この『属性』にこだわりがある方は, レビューなどを確認の上ご利用下さい」, 「この『属性』の評価はアイテムの評価に大きな影響を与えています」等の表現が見られ

また各リフト値がとる値にあまり差が見られない場合、「バランスのとれたアイテムです」、「各要素よりもアイテム全体を見て評価されている」等の表現がされていた。図 4.8(a) の 2 件目の説明文がこれに当てはまる。この場合も、 $(Y_{ij}^a \cap X_i^T)$ を満たすレビューの数から判断している回答も存在した。このようにアイテムモデルにおいてリフト値が大きな特徴を持たない場合でも、説明文を作成可能であることがわかる。

- ・このホテルは客室、接客対応、コストパフォーマンスに優れており、これらを重視する方にとってもおすすめです！
(僅かではあるが、客室、コスパ、接客対応のAのリフト値が高く、また、客室、コスパ、接客対応の属性評価：好評の合計数がとても多いため)
- ・全体的にバランスのとれたホテルである。
(Aのリフト値が、属性全7つあるうち5つが1以上のため突出した特徴はないと判断。)
- ・ですが、客室やお風呂へのこだわりが強い方は、写真やレビューをよく確認してください！
客室やお風呂が気に入らないようなら残念ながら、お勧めできません。
(客室と風呂のDのリフト値がとても高いため)

(a) アイテム1の回答例

- ・部屋の快適さをおすすめできます。(リフト値A：客室が高いため)
- ・コスパ・客室の2つの要素で特に優れているホテルと言える。
(Aのリフト値でコスパと客室が高いため)
- ・このホテルは客室とコストパフォーマンスに特に優れています！
これらを重視する方には強くおすすめできるホテルです。
(コスパと客室のAのリフト値が高いため)

(b) アイテム2の回答例

- ・この映画は、物語を気に入るかどうかで好き嫌いが分かれる。
(物語のリフト値でAとDが高いため)
- ・この作品は主に物語が評価されている(物語のリフト値Aが高い)
- ・この映画は映像に関する評価が高いため映像面に期待する人は好きかも知れない。
(映像のAのリフト値が1を越えていることと、映像のAの合計数が大きいため)

(c) アイテム5の回答例

図 4.8: 実際にされたアンケートの回答例の一部

4.1.3 推薦システムでの利用に関する考察

アイテムモデルを利用した推薦説明について、推薦システムでの利用方法を考察する。実験結果を踏まえ、4種類のリフト値それぞれを根拠とした説明文として、表4.2に挙げるものが利用可能と考える。

表 4.2: 閾値を超えたリフト値に応じて提示する説明文の例

リフト値の種類	提示する文章の例
$lift_{ij}$ ($pos \Rightarrow pos$)	<ul style="list-style-type: none"> この「属性」に興味(こだわり)を持つ方には、「アイテム」を気に入るでしょう。 この「属性」の評価は、「アイテム」の評価に大きな影響を与えています。 この「属性」の評判は良いです。
$lift_{ij}$ ($pos \Rightarrow neg$)	<ul style="list-style-type: none"> この「属性」にこだわりを持たない方は、他の要素で判断するほうが良いかもしれません。 この「属性」の評価が、「アイテム」の評判に繋がることはありません。 この「属性」に興味(こだわり)を持たない方でも、「アイテム」を気に入るでしょう。
$lift_{ij}$ ($neg \Rightarrow pos$)	<ul style="list-style-type: none"> この「属性」の評価が、「アイテム」の評判に繋がることはありません。 この「属性」に興味(こだわり)を持たない方でも、「アイテム」を気に入るでしょう。
$lift_{ij}$ ($neg \Rightarrow neg$)	<ul style="list-style-type: none"> この「属性」に興味(こだわり)を持つ方は、事前にそれに関するレビューに目を通すことをおすすめします。 この「属性」の評価は、「アイテム」の評価に大きな影響を与えています。

推薦システムでは表4.2に示したような文書パターンをテンプレートとして用意しておき、各リフト値が設定した閾値を超えた場合、該当するテンプレートを用いて説明文を生成できる。リフト値に大きな特徴が見られない場合は、 $(Y_{ij}^a \cap X_i^r)$ を満たすレビューの数による判断を行うか、「バランスのとれた(平均的に見られている)アイテムです」等の説明を行うことが可能である。

図4.9に推薦システムでの利用を想定した説明文の生成例を示す。例えばリフト

値 ($neg \Rightarrow neg$) の閾値を 1.5 と設定した場合、閲覧アイテムが図 4.9(a) のようなアイテムモデルを持っていれば、対象ユーザには図 4.9(b) や図 4.9(c) のような説明文が提示される。図 4.9(b) は図 3.7 の提示方法に基づき、図 4.9(c) は図 3.8 の提示方法に基づいている。

属性	$pos \Rightarrow pos$
物語	1.74
配役	0.97
演出	1.55
映像	1.02
音楽	0.93

(a) ある映画のアイテムモデル

この映画の特徴

この映画の物語、演出を気に入って頂けた場合、映画自体を気に入って頂けるでしょう。

(b) 提示される説明文その 1

この映画の特徴

ここを気に入った人は映画も好きになれるでしょう。

物語

演出

(c) 提示される説明文その 2

図 4.9: 推薦システムを想定した、アイテムモデルからの説明文生成と提示例

4.2 評価実験

提案するアイテムモデルの有効性を検証するために実施した評価実験に関して説明する。始めに実験概要を説明し、次に実験で使用するデータセット、システムにおける推薦説明と推薦アルゴリズムの詳細を述べた後、実験結果について考察を行う。

4.2.1 実験概要

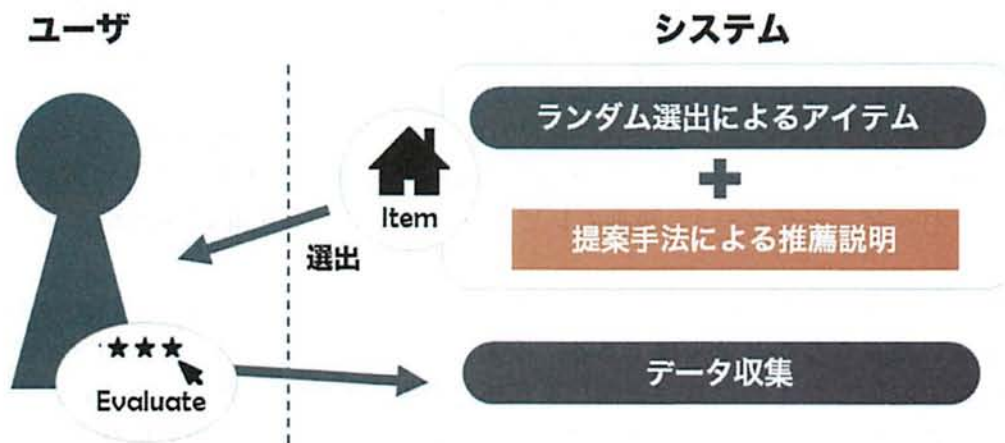
本実験の目的は、提案手法を用いた推薦システムを協力者に利用してもらい、推薦アルゴリズムに対する精度、アイテム評価時に参考にした推薦説明の内容を分析することで、提案手法の有効性を検証することである。

実験では、工学系大学生および大学院生 23 名に対し、構築した推薦システムによりアイテムを評価してもらうことで、その結果を収集した。アイテムには宿泊施設のデータセットを使用した。データセットの詳細は 4.2.2 項にて説明する。協力者には「友達、または家族との旅行」という利用目的のもと、宿泊施設を利用したいかどうか評価してもらった。また設定した利用目的に基づき一貫した評価をしてもらうため、実験はなるべく期間をあげず、集中して行ってもらった。

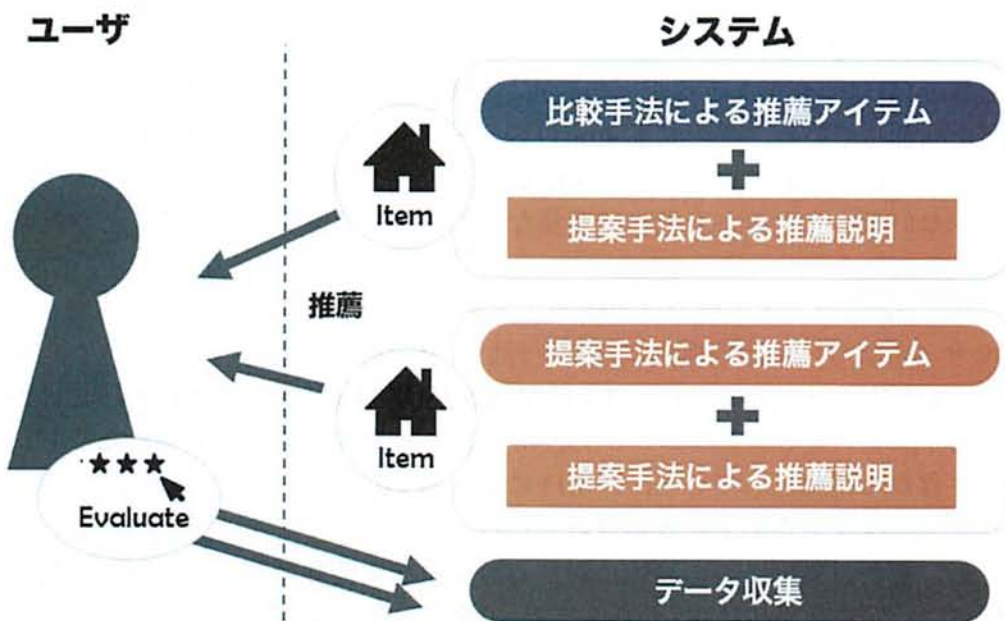
実験の概要を図 4.10 に示す。始めに図 4.10(a) に示すように、協力者にはシステムがランダムに選んだアイテムを評価してもらった。次に図 4.10(b) に示すように、協力者には比較手法による推薦アイテムと、提案手法による推薦アイテムをそれぞれ評価してもらった。比較手法にはアイテムベース推薦の代表的なアルゴリズムである Sarwar ら [15] の手法を採用し、提案手法には 3.3 節で提案した価値観アイテムモデリングによる推薦を用いた。

アイテムの評価時には、アイテムの基本情報に加えて提案手法に基づいた推薦説明を提示し、その情報をもとにアイテムに評価値を下してもらった。また評価値の付与にあたって参考になった推薦説明も選択してもらった。具体的なアイテムの提示内容と評価方法については、4.2.3 項と 4.2.5 項にて説明する。

収集したアイテムの評価値により、比較手法と提案手法の推薦精度を比較する。また参考になった推薦説明を分析することで、提案手法による推薦説明の有効性を検証する。



(a) 行程その 1 : テストアイテムの評価



(b) 行程その 2 : 推薦アイテムの評価

図 4.10: 評価実験の概要図

4.2.2 システムで利用するデータセット

データセットの選択

協調フィルタリングにおける評価実験の代表的なデータセットは MovieLens が挙げられるが、MovieLens では属性評価値を持たないため、提案手法に用いることはできない。アイテムの総合評価値以外にその属性評価値も付与されている大規模なデータセットとしては、宿泊施設のレビューサイトである楽天トラベル¹や 4travel²、家電のレビューサイトである価格.com³、美容院の予約サイトであるホットペッパービューティー⁴、映画のレビューサイトである Yahoo!映画などが存在する。

アイテムの評価を想定した場合、どのデータセットもアイテムを実際に使用して評価してもらうことは困難である。検索エンジンなどを利用してそのアイテムの情報を収集し、これに基づき価値を判断してもらう方法も考えられるが、本実験の目的である、アイテム評価時における参考になった推薦説明の検証が困難になると考える。そこでこれらを考慮して、協力者にはシステム内で提示した情報のみからアイテムを評価してもらうことにした。その場合、美容院や映画などのアイテムは、提示されたアイテムの情報のみからその良し悪しを判断することは特に難しいと考えられる。

三澤ら [4] は、楽天トラベル、4travel、価格.com におけるデータセットの評価値分布を検証している。その結果 4travel のデータセットが一番評価値の分布に偏りが無いことが示されている。また 4travel では属性の評価項目が 7 つと多く、属性数が少ないデータセットと比べて多様なアイテムモデルが得られることが期待できる。

これらを踏まえ、本実験では 4travel のデータセットを利用した。データセットは公式 サイトより、都道府県別に検索を行い、その検索結果のページをスクレイピングして収集したものである。収集期間は 2013 年 11 月 2 日から 11 月 15 日にかけてであり、データセットはユーザ 15,948 名、アイテム 11,825 件からなるレビュー 83,452 件から構築されている。図 4.11 に 4travel の評価データの例を示す。図 4.11 のレビュータイトルの下に表示されている、アイテムに対する評価値を総合評価値とし、それ以外の 7 つの属性に対する評価値を属性評価値として扱う。総合評価値および属性評価値は 0.5 から 5 まで、0.5 刻みの 10 段階で評価を行っている。

¹<http://travel.rakuten.co.jp/>

²<http://4travel.jp/>

³<http://kakaku.com/>

⁴<http://beauty.hotpepper.jp/>

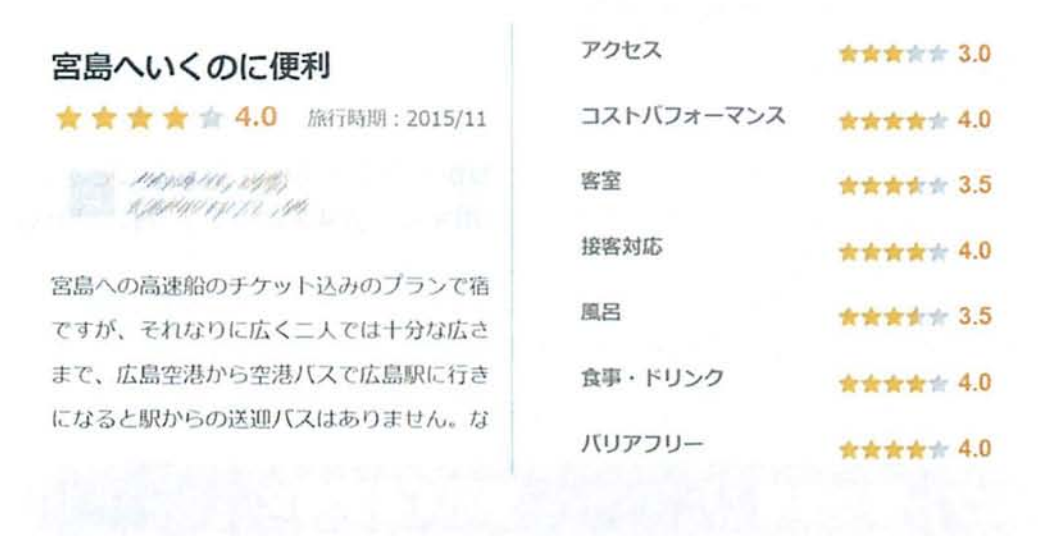


図 4.11: 4travel の評価データの例

データセットの加工

投稿されたレビュー数の関係からアイテムによっては、図 3.5 で示したように各属性における 4 種類のリフト値が全て求められるわけではない。この場合もアイテム間の類似度は計算可能であるが、推薦説明として利用できるリフト値が少なくなる可能性がある。構築するシステムは、実際の利用を想定したものではなく、提案手法の有効性を検証する目的であるため、アイテムへの推薦説明の提示を確実に行う方が好ましい。そこでデータセットの全アイテムについてアイテムモデルを構築し、各属性における 4 種類のリフト値が全て算出されたアイテムのみを実験に使用した。加工後のデータセットにおけるアイテムは 1,394 件である。

4.2.3 システムにおける推薦説明の提示方法

推薦説明の提示内容

プロトタイプシステムにおける、推薦説明の提示方法について説明する。システムではアイテムの選択時に、ユーザによりわかりやすく情報を伝えることを考慮して、図 3.8 の方法に基づいた提示方法を採用する。表 4.3 に示すように、事前に各 4 種類のリフト値が大きい場合に対応した説明文をそれぞれ用意しておき、各説明文において閾値を超えたリフト値を持つ属性のアイコン画像を提示するようにする。推薦説明で提示する説明文は表 4.2 に基づいて作成した。

表 4.3: システムで提示する価値観モデリングに基づいた説明文

リフト値の種類	提示する説明文
$lift_{ij}(pos \rightarrow pos)$	ここの要素にこだわりや興味を持つは宿泊施設を気に入るでしょう
$lift_{ij}(pos \rightarrow neg)$	ここの要素にこだわりや興味を持った人でも宿泊施設を気に入るでしょう
$lift_{ij}(neg \rightarrow pos)$	ここの要素にこだわりや興味を持った人でも宿泊施設を気に入るでしょう
$lift_{ij}(neg \rightarrow neg)$	ここの要素は、宿泊施設の評価に大きな影響を与えています

また価値観アイテムモデルとは別に、アイテムの持つ基本情報に基づいた推薦説明の提示も行った。表 4.4 に示すように、アイテムの各属性評価値の平均値が好評、不評、それ以外の 3 つの場合の説明文を用意しておき、それぞれに該当する平均値評価値を持つ属性を提示する。

表 4.4: システムで提示する基本情報に基づいた説明文

属性評価値の分類	提示する説明文
好評	ここの要素は、好評です
不評	ここの要素はあまり評判が良くないかもしれません
それ以外	ここの要素は、平均的に見られています

このように価値観アイテムモデルだけでなく、基本情報に基づいた推薦説明も提示し、その結果を比較することで提案手法の有用性を検証することを試みる。

価値観アイテムモデリングを使用した推薦説明における閾値に関しては、提示される推薦説明の内容に偏りが無いことが望ましいと考える。事前調査において、全リフト値について同一の閾値を用いた場合、各アイテムにおいて推薦説明として検出されるリフト値の属性の数に大きな偏りが生じることを確認した。例えば $lift_{ij}(neg \Rightarrow neg)$ では、多数の属性に対し説明文が生成されるのに対し、 $lift_{ij}(pos \Rightarrow neg)$ に関しては属性が全く検出されずに、推薦説明が作成できないことが多くみられた。このように提示する推薦説明の内容に大きな偏りが出してしまうと、推薦説明に対するユーザの反応を適確に検証することが困難になる可能性がある。

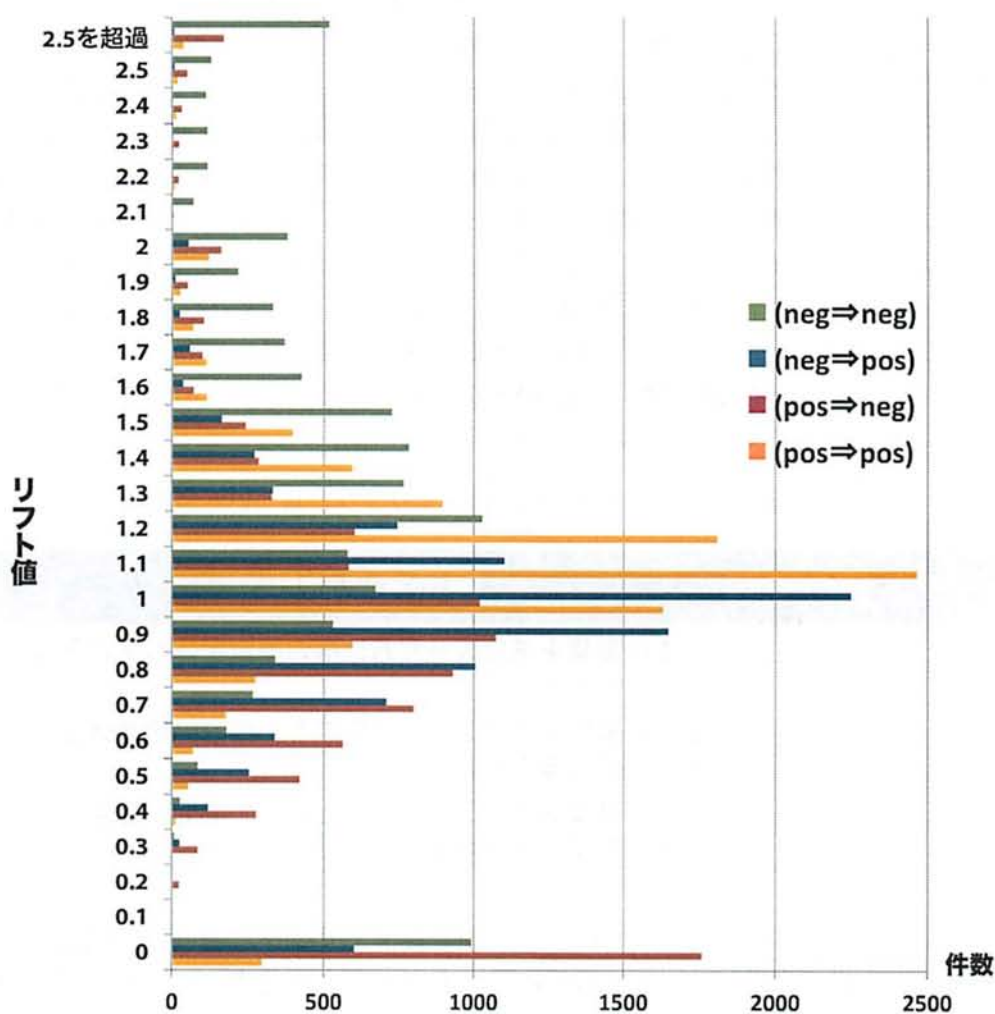


図 4.12: データセットのアイテムにおけるリフト値の分布

データセットの全アイテム 1,394 件における、各種類のリフト値の分布を図 4.12 に示す。図 4.12 から、 $lift_{ij}(pos \Rightarrow neg)$ のリフト値は 1 以下に偏って分布している

ことが分かる．また 4.1 節で行った予備実験では、 $lift_{ij}(pos \Rightarrow neg)$ の解釈が難しいとの意見が協力者から多く寄せられた．そこで評価実験では $lift_{ij}(pos \Rightarrow neg)$ に基づいた説明文の提示を行わず、それ以外の $lift_{ij}(pos \Rightarrow pos)$ 、 $lift_{ij}(neg \Rightarrow pos)$ 、 $lift_{ij}(neg \Rightarrow neg)$ については各リフト値毎に異なる閾値を設定し、説明文が提示される回数に大きな偏りが出ないようにした．

説明文の修正

評価実験に用いるシステムを用いたテストプレイの中で、協力者からは表 4.3 で提示された説明文について、その解釈が困難である、意味がわかりにくい、などと言った意見が多く寄せられた．特に「興味、こだわりを持つ」という条件や、「属性の評価がアイテムの評価に影響を与える、左右される」といった表現が難しいとの指摘が多かった．推薦説明における説明文は、表 4.2 の予備実験の結果に基づいたものである．予備実験では売り手が買い手側を意識して表現を作成してもらっており、実験協力者はリフト値や価値観モデルに関し理解した上で説明文を生成していた．これに対し、今回の実験では実験協力者は買い手の立場であり、リフト値や価値観モデルについての知識を特に持っていない状態である．このような、アイテムに対する売り手と買い手の知識の差が、説明文がわかりにくかった原因と考える．そこで本実験では、予備実験の結果を考慮しつつ、協力者が買い手の場合でもわかりやすい表現に説明文を修正した．

表 4.5: 評価実験で利用する価値観モデリングに基づいた説明文

リフト値の種類	提示する説明文
$lift_{ij}(pos \rightarrow pos)$	ここの要素を気に入った方は宿泊施設全体も気に入って頂けるでしょう
$lift_{ij}(neg \rightarrow pos)$	ここの要素を気に入らない方でも、宿泊施設全体は気に入って頂けるでしょう
$lift_{ij}(neg \rightarrow neg)$	ここの要素を気に入らなかった方は、宿泊施設全体も気に入って頂けない可能性があります

表 4.5 に、最終的にシステムで提示した価値観モデリングに基づいた説明文を示す．表 4.5 の説明文は、表 4.1 の結果と買い手側の視点を考慮して再度作成されたものである．表 4.1 における (1)、(4) の説明文は表 4.4 で使用するため、リフト値に基づいた推薦説明には使用しないことにした．また協力者の意見をもとに「こだわり、アイテムへの影響」などの表現も使わないようにした． $lift_{ij}(pos \Rightarrow pos)$ が大きい場合には、(2) の説明文が多く存在したので、(2) に基づき説明文を生成した． $lift_{ij}(neg \Rightarrow pos)$ では、使用しない説明文を除いた場合、どの種類の説明文も

同程度に存在していたので、一番リフト値の意味に即した (9) に基づき説明文を生成した。 $lift_{ij}(neg \Rightarrow neg)$ についても、 $lift_{ij}(neg \Rightarrow pos)$ と同じ理由から (5) に基づき説明文を生成した。

4.2.4 システムで利用する推薦アルゴリズム

本項では評価実験にあたって、価値観モデルのアイテムベース推薦への適用方法について説明する。3.3 節で述べた、提案手法による以下の3つのアイテム間類似度の計算方法について、その特性を分析した。 m はアイテムの持つ属性数を意味している。

1. 全てのリフト値を1つの $4m$ 次元ベクトルにまとめ、そのベクトル間の類似度を計算する方法
2. p_a, p_t (3.2.3 項参照) の組み合わせによる4種類のリフト値を、各属性毎に4次元のベクトルとしてまとめ、それぞれの類似度を計算する方法
3. 全 m 個の属性から計算されるリフト値を、 p_a, p_t の組み合わせ別に m 次元のベクトルとしてまとめ、それぞれ類似度を計算する方法

4.2.2 項で加工したデータセットの全アイテム同士における、(1) の手法を用いた Pearson 相関係数を算出し、その結果をヒストグラムにして図 4.13 に示す。図 4.13 における縦軸は、計算された全アイテム間の Pearson 相関係数 ($1,394 \times 1,394$) のうちに占める割合を示している。手法 (2), (3) に関しても、同様に求めた Pearson 相関

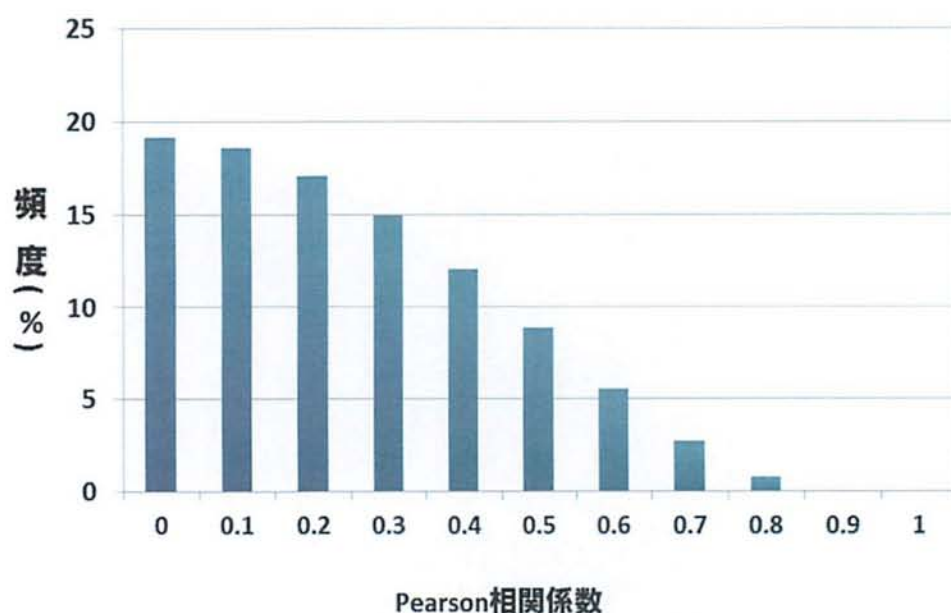


図 4.13: 手法 (1) を用いた Pearson 相関係数の分布図

係数のヒストグラムにしたものを図 4.14, 図 4.15 にそれぞれ示す。図 4.14, 図 4.15 では各属性毎、または各 p_a, p_t の組み合わせ毎に類似度が計算されるので、データ

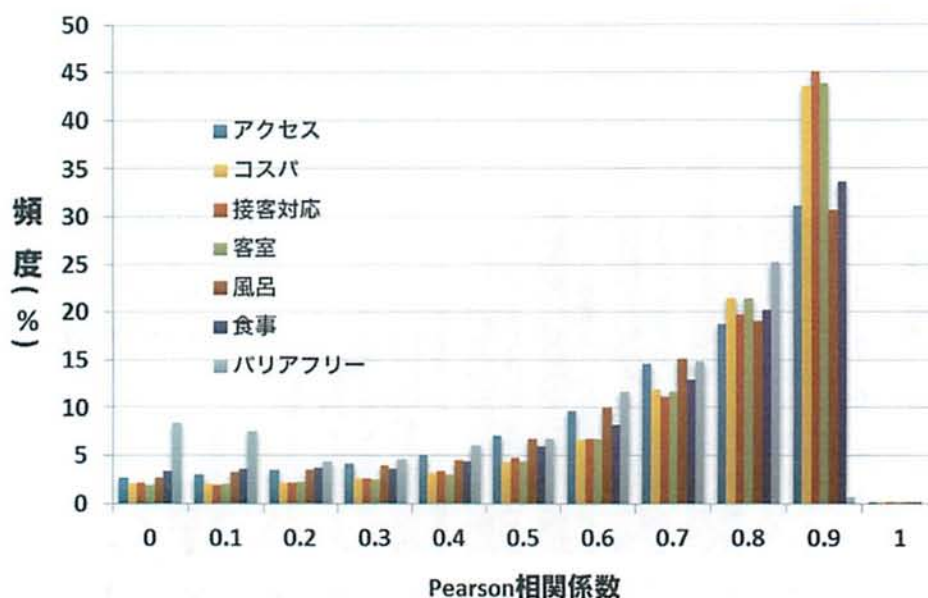


図 4.14: 手法 (2) を用いた Pearson 相関係数の分布図

の系列を複数にして載せている。図 4.13, 図 4.14, 図 4.15 より, 手法 (1), (3) では Pearson 相関係数の分布が 0 に偏っている傾向にあることがわかる。一方で手法 (2) では Pearson 相関係数の分布が 1 に偏っていることがわかる。これは比較対象であるベクトルの次元数が手法 (1) では 28 次元, 手法 (3) では 7 次元なのに対し, 手法 (2) では 4 次元と少ないため, 類似度が適確に計算されない場合が多くなったと考察できる。手法 (3) では類似アイテムの, さらに属性に対する共通の見られ方を推論できるが, 評価実験ではその推論は必要ないため, 手法 (1) による類似度計算が最適であると考ええる。

4.2.3 項で述べたように, 推薦説明では $lift_{ij}(pos \Rightarrow neg)$ のリフト値を利用した説明文の提示を行わないことにした。手法 (1) において $lift_{ij}(pos \Rightarrow neg)$ を除いた 21 次元ベクトルで類似度計算をした方法を手法 (4) とした場合, 手法 (4) における全アイテム間の Pearson 相関係数の分布図をヒストグラムにて図 4.16 に示す。

図 4.16 より $lift_{ij}(pos \Rightarrow neg)$ を除いても, Pearson 相関係数の分布が 0 に偏っており, 類似度の計算に大きな影響はないと考えられる。これらを踏まえ, 評価実験では手法 (4) を適用して価値観モデリングに基づいたアイテム間の類似度計算を行う。

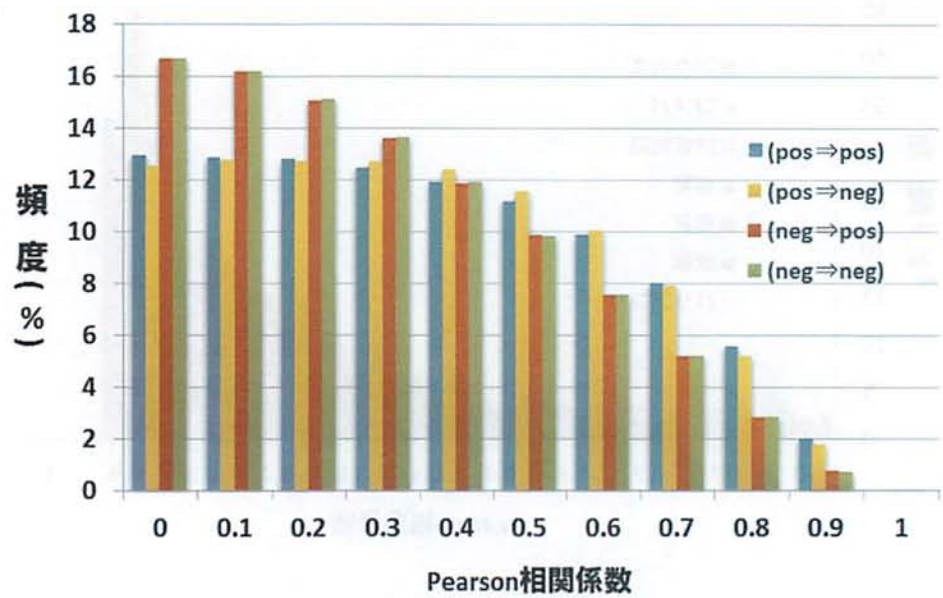


図 4.15: 手法 (3) を用いた Pearson 相関係数の分布図

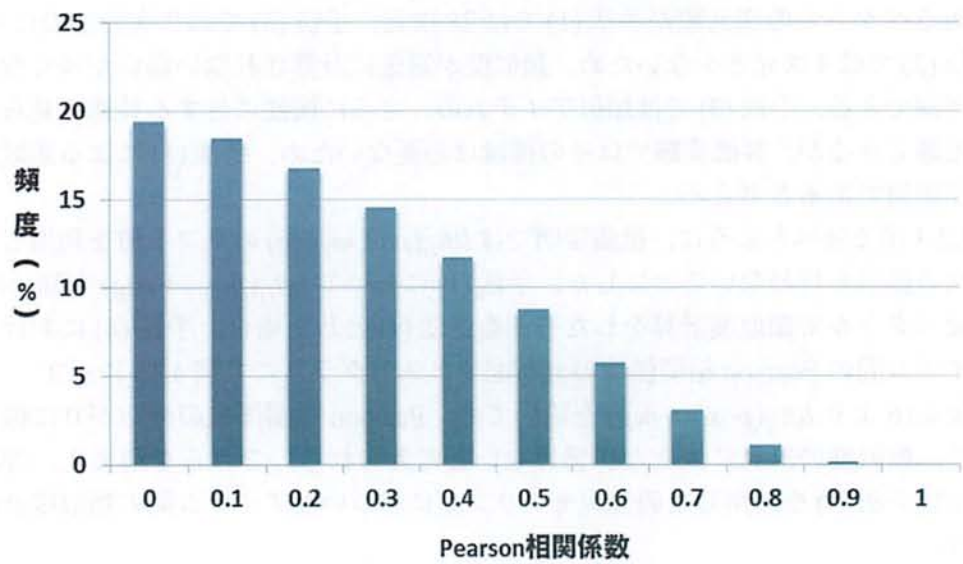


図 4.16: 手法 (4) における Pearson 相関の分布図

4.2.5 プロトタイプシステムとデータの収集方法

本項では構築したプロトタイプシステムを利用した実験手順について説明する。プロトタイプシステムは ruby on rails 4 を用いてサーバ上に構築し、オンラインで実験に参加できるようにした。

実験は、テストアイテムの評価と推薦アイテムの評価から構成される。始めに協力者は、図 4.17 に示すようにデータセットからランダムに選択した 15 件の宿泊施設をテストアイテムとして評価する。



図 4.17: テストアイテムの一覧

アイテムの選択時には、図 4.18 の宿泊施設の基本情報、図 4.19 の推薦説明、図 4.20 の評価欄が提示される。図 4.19 では表 4.4 と表 4.5 に基づき、属性評価の平均値と、価値観アイテムモデルをもとにした 6 つの説明文を提示している。協力者は図 4.18、図 4.19 の提示内容を参考に、設定した目的に従って施設を利用したいかどうかを評価してもらうと同時に、付与した評価に対する理由も回答する。評価はデータセット元の 4travel と統一して、0.5 から 5 までの 0.5 刻みによる 10 段

階評価にした。

評価値の付与と平行して、協力者には評価において参考にした推薦説明を、図 4.21 に示すように選択してもらう。推薦説明が参考になった場合、その提示内容が施設を利用する価値を上げる観点から参考になった場合は「参考になった(+情報)」，利用価値を下げる意味で参考になった場合は「参考になった(-情報)」を選択する。またどちらの「参考になった」を選択しても、その理由を回答してもらう。図 4.21 では協力者がお風呂を重視しておらず、提示された推薦説明が施設の利用価値を上げる情報であったため「参考になった(+情報)」を選択し、その理由を記入している例を示している。これらの評価内容を、協力者は 15 個のテストアイテム全てに行う。

次に協力者は、2.1.2 項で述べた比較手法により推薦された宿泊施設を 10 件、提案手法により推薦された宿泊施設を 10 件、合計 20 件の推薦アイテムについて評価を行う。比較手法、提案手法による推薦アイテムはわけて提示せず、図 4.17 と同様の形式で混在させて表示することで、協力者が推薦手法によって恣意的に評価内容を偏らせることができないようにする。推薦アイテムの評価方法は、テストアイテムに対する場合と同じである。ただしユーザの負荷を考慮して、推薦アイテムの評価では参考になった場合の理由の記入(図 4.21)は省いた。

協力者は 20 個の推薦アイテムを評価した後、最後にアンケートに回答する。アンケートでは、協力者の基本情報、施設の目的、評価時に重視した属性などに回答する。

HOME

進行度の確認

(1) 実験マニュアル

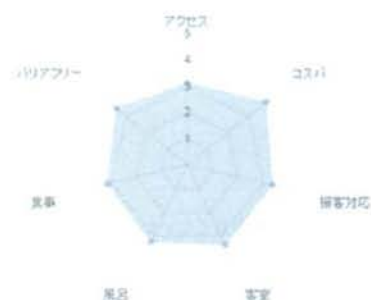
(2) ステップAを行う

ホテルアルファワン能登和倉 [スタンダードホテル]

宿泊施設の基本情報



宿泊施設の簡単な説明: なし



(施設全体の総合評価 3.29/5.0)

住所: 石川県七尾市石崎町香島1-20

アクセス: JR和倉温泉駅より車で3分

宿泊施設の設備:

- | | | | |
|---|--|---|--|
| <input checked="" type="checkbox"/> 温泉 | <input checked="" type="checkbox"/> 大浴場 | <input checked="" type="checkbox"/> 露天風呂 | <input checked="" type="checkbox"/> 貸切風呂 |
| <input checked="" type="checkbox"/> 源泉掛け流し | <input checked="" type="checkbox"/> 混浴あり | <input checked="" type="checkbox"/> サウナ | <input type="checkbox"/> インターネット可 |
| <input checked="" type="checkbox"/> ジム/フィットネス | <input checked="" type="checkbox"/> プール | <input type="checkbox"/> 駐車場 (有料+無料) | <input type="checkbox"/> 洗浄機能トイレ |
| <input checked="" type="checkbox"/> エステ | <input type="checkbox"/> マッサージ | <input checked="" type="checkbox"/> ペット対応 | <input checked="" type="checkbox"/> 送迎サービス |
| <input type="checkbox"/> ランドリーサービス | <input checked="" type="checkbox"/> コンビニ近く | <input checked="" type="checkbox"/> 駅から5分以内 | <input checked="" type="checkbox"/> 航空券・JR付き |
| <input type="checkbox"/> パソコン貸出し | | | |

図 4.18: アイテム選択時に提示する基本情報

宿泊施設の特徴

<p>《この要素は好評です》</p> <p>アクセス コスパ 接客対応 バリアフリー</p> <p>参考にならなかった(評価に影響を与えなかった) ▼</p> <p>*3つの中からどれかを選択</p> <p>その理由↓</p> <p>*「参考になった」を選択した場合、随時で良いので必ず記入</p> <p></p>	<p>《この要素は平均的に見られています》</p> <p>風呂 食事</p> <p>参考にならなかった(評価に影響を与えなかった) ▼</p> <p>*他5箇所についても同様</p> <p>その理由↓*他5箇所についても同様</p> <p></p>
<p>《この要素はあまり評判が良くないかもしれません》</p> <p>客室</p> <p>参考にならなかった(評価に影響を与えなかった) ▼</p> <p>その理由↓*</p> <p></p>	<p>《この要素を気に入った方は、宿泊施設全体も気に入って頂けるでしょう》</p> <p>アクセス 接客対応 食事</p> <p>参考にならなかった(評価に影響を与えなかった) ▼</p> <p>その理由↓*</p> <p></p>
<p>《この要素を気に入らなかった方は、宿泊施設全体も気に入って頂けない可能性があります》</p> <p>アクセス コスパ 接客対応 食事</p> <p>参考にならなかった(評価に影響を与えなかった) ▼</p> <p>その理由↓*</p> <p></p>	<p>《この要素を気に入らない方でも、宿泊施設全体は気に入って頂けるでしょう》</p> <p>風呂</p> <p>参考にならなかった(評価に影響を与えなかった) ▼</p> <p>その理由↓*</p> <p></p>

図 4.19: アイテム選択時に提示する推薦説明

宿泊施設を評価する

この宿泊施設に対する評価（『必須』）

3 ▼

またその評価値にした理由を簡潔で良いので書き下さい（『必須』）

その他何かコメントがあれば書き下さい

図 4.20: アイテム選択時に提示する評価欄

全体

《ここの要素を気に入らない方でも、宿泊施設全体は気に入って頂けるでしょう》

風呂

参考になった(+情報、評価を上げる情報として役立った) ▼

参考にならなかった(評価に影響を与えなかった)

参考になった(+情報、評価を上げる情報として役立った)

参考になった(-情報、評価を下げる情報として役立った)

お風呂は特に重視していないので。

図 4.21: 推薦説明が参考になった場合の選択とその理由の記入例

4.2.6 評価実験の結果と考察

本項では評価実験で得られた結果をまとめ、考察を行う。始めにアンケートの回答結果を示し、次に推薦精度と推薦理由の効果について説明する。

アンケートでは協力者に、評価実験で重視した属性を回答(複数可)してもらった。その結果を図 4.22 に提示する。また図 4.23 に協力者が実験に要した時間を示す。図 4.22 より、接客対応、バリアフリーを重視する人は少なかった。これは協

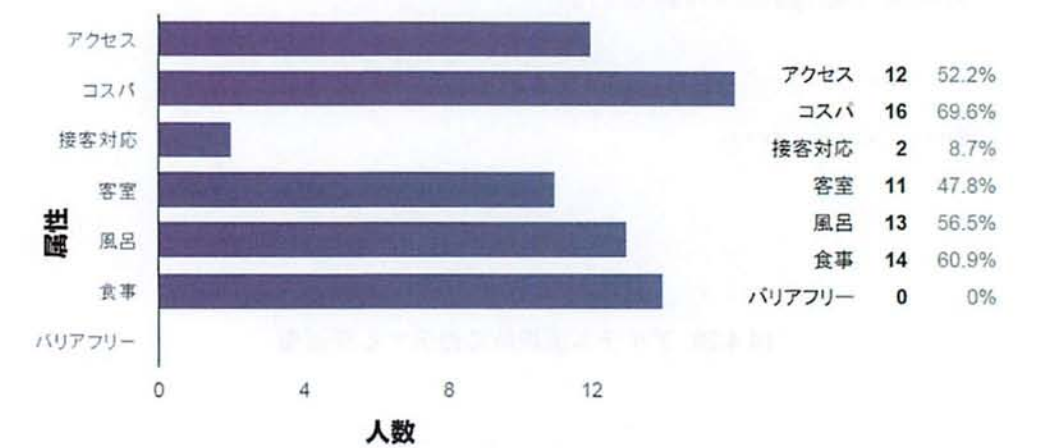


図 4.22: 評価実験で重視した属性

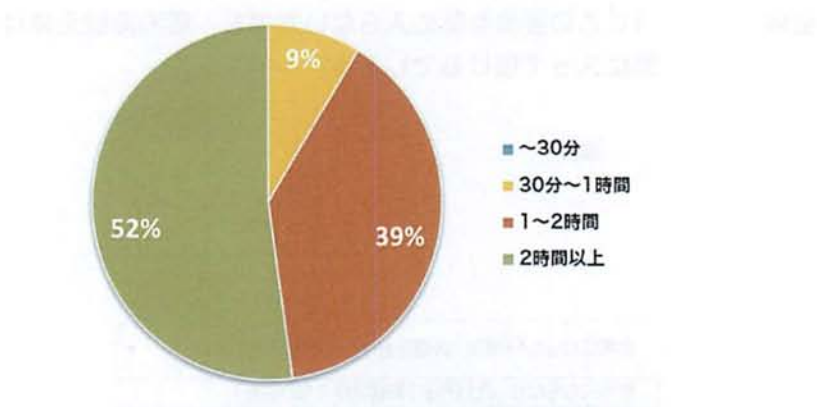


図 4.23: 評価実験に要した時間

力者が学生であり、施設の利用目的の多くが旅行を中心としたものであったことが原因と考える。また図 4.23 より、評価実験に 2 時間以上を要した人が多くみられた。これは提案手法の評価のために、協力者に参考になった推薦説明の選択や評価理由の記述をしてもらったことが原因と考える。

23 名の協力者から、テストアイテム 345 件、比較手法による推薦アイテム 230 件、提案手法による推薦アイテム 230 件の合計 805 件の評価データが得られた。表

4.6 に比較手法と提案手法の推薦精度を示し、表 4.7 に各協力者の比較手法、提案手法における平均評価値を示す。表 4.7 の 4 列目では比較手法と比べて提案手法の平均評価値が大きい場合に ○ と表示している。例えば 1 行目は、協力者 1 について、比較手法の推薦アイテム 10 個に対する平均評価値が 3.05、提案手法の場合が 3.60 であり、提案手法の方が平均評価値が大きいことを意味している。また表 4.7 の各協力者の平均評価値を対応のある標本として、比較手法と提案手法の t 検定を行った結果を表 4.8 に示す

表 4.6: 比較手法と提案手法の精度

	比較手法	提案手法
精度	3.113043	3.280435

表 4.6 より比較手法と比べ、提案手法の精度の方が高くなっている。両側検定で $p > .05$ であったため有意差は確認できなかったが、 p 値は 0.075 と 0.05 に近く、また表 4.7 より協力者 23 人のうち 15 人は提案手法の方が平均値評価値が大きいことがわかる。

全アイテム 805 件について、評価時に参考になった説明文の数を図 4.24、図 4.25 に示す。図 4.24、図 4.25 の縦軸は全 805 件における参考になった説明文の数の割合を意味し、「+情報」として参考になった結果を青色、「-情報」として参考になった結果を赤色、その合計を黒色で示している。図 4.24 では属性の平均値評価に基づいた説明文について参考になった数を載せており、横軸の好評、平均、不評は表 4.4 の好評、それ以外、不評に対応している。図 4.25 では価値観モデルに基づいた説明文について参考になった数を載せており、横軸の $(pos \Rightarrow pos)$ 、 $(neg \Rightarrow pos)$ 、 $(neg \Rightarrow neg)$ は表 4.5 の $lift_{ij}(pos \Rightarrow pos)$ 、 $lift_{ij}(neg \Rightarrow pos)$ 、 $lift_{ij}(neg \Rightarrow neg)$ にそれぞれ対応している。

属性の平均値評価に基づいた推薦説明では、図 4.24 より好評と不評に関する説明文を参考にした人が多く見られた。また好評の説明文を参考にした人は主に「+情報」として、不評の説明文を参考した人は主に「-情報」として解釈しており、好評、不評の情報に対するユーザの判断として妥当性のある結果であると考ええる。

図 4.25 より、 $lift_{ij}(pos \Rightarrow pos)$ 、 $lift_{ij}(neg \Rightarrow pos)$ 、 $lift_{ij}(neg \Rightarrow neg)$ の説明文は、どれも全体のうち 2 割から 3 割程の比率で参考になっていることがわかる。このことから、価値観モデリングに基づいた推薦説明は、アイテムを評価する上でユーザの意思決定に影響を与える情報であったと考える。

$lift_{ij}(pos \Rightarrow pos)$ では、図 4.25 から「+情報」として参考になった意見が多いことがわかる。「+情報」として参考になった人の理由をみると、多くは提示された説明

表 4.7: 各協力者毎の平均評価値

ユーザID	比較手法	提案手法	提案手法> 比較手法?
1	3.05	3.60	○
2	2.55	3.00	○
3	3.15	2.70	
4	3.40	3.85	○
5	2.75	3.05	○
6	3.25	3.05	
7	3.00	3.15	○
8	3.60	4.20	○
9	2.85	3.40	○
10	3.70	3.60	
11	3.35	3.85	○
12	3.00	3.10	○
13	3.10	3.60	○
14	3.90	3.35	
15	4.05	3.55	
16	2.75	2.85	○
17	2.95	3.55	○
18	2.75	2.90	○
19	3.15	2.65	
20	2.05	2.90	○
21	2.70	3.35	○
22	3.30	2.95	
23	3.25	3.25	

文に含まれる属性を重要視していた。一方で少数ではあるが、 $lift_{ij}(pos \Rightarrow pos)$ の説明文を「-情報」として参考にした意見も存在した。この場合、提示された属性の中に、重要視している属性が存在していなかったため消極的に捉えていた。このことから $lift_{ij}(pos \Rightarrow pos)$ は、提示された説明文に含まれる属性を重要視する人にとっては、肯定的な情報として役に立つことが考察できる。

$lift_{ij}(neg \Rightarrow neg)$ では図 4.25 より、主に消極的な情報として参考になっていることがわかる。その時の協力者の参考となった理由は、提示された属性を重要視していたため、という意見が多くみられた。また少数の意見として、提示された属性に重要視していた属性が含まれなかったため「+情報」として参考になったという意見も見られた。このことから $lift_{ij}(pos \Rightarrow pos)$ に基づいた説明文は、提示され

表 4.8: 比較手法と提案手法の t 検定結果

(a) 対応サンプルの統計量

	比較手法	提案手法
統計量	23	23
平均値	3.113043478	3.280434783
標準偏差	0.201413043	0.160167984

(b) 対応サンプルの t 検定結果

比較手法と提案手法	
平均値の差	-0.16739
ピアソン相関	0.492446884
仮説平均との差異	0
自由度	22
t 値	-1.868023851
$P(T \leq t)$ 片側	0.037568174
t 境界値 片側	1.717144374
$P(T \leq t)$ 両側	0.075136349
t 境界値 両側	2.073873068

た説明文に含まれる属性を重要視する人にとっては、アイテムの注意を喚起する情報として役に立つことが考察できる。

$lift_{ij}(neg \Rightarrow pos)$ が参考になった数は、図 4.25 から他のリフト値に基づいた説明文に比べると少ないことがわかる。また $lift_{ij}(neg \Rightarrow pos)$ の説明文は、肯定的、否定的として解釈された割合がどちらも同じ程度だとわかる。参考になった理由をみると、「+情報」として参考にした場合、提示された属性を重要視しているからこそ利用価値が上がった意見があれば、重要視していないので利用価値が上がったとの意見もあった。また「-情報」についても、提示された属性を重要視しているからこそ利用価値が下がったという意見や、重要視していないので利用価値が下がったという意見があった。このことから $lift_{ij}(neg \Rightarrow pos)$ の捉え方は人や状況に依存し、その解釈が困難であったため、推薦説明として参考になった数が少なかったと考察できる。

属性が好評であることに基づいた説明文では、「-情報」として役に立った意見が全体の約 2% しかないのに対し、 $lift_{ij}(pos \Rightarrow pos)$ では約 16% であり、反対として役に立った意見の割合が多いことがわかる。これは $lift_{ij}(pos \Rightarrow pos)$ が「この『属性』を気に入る人は」という条件を用いた説明文であり、当該属性が好評でな

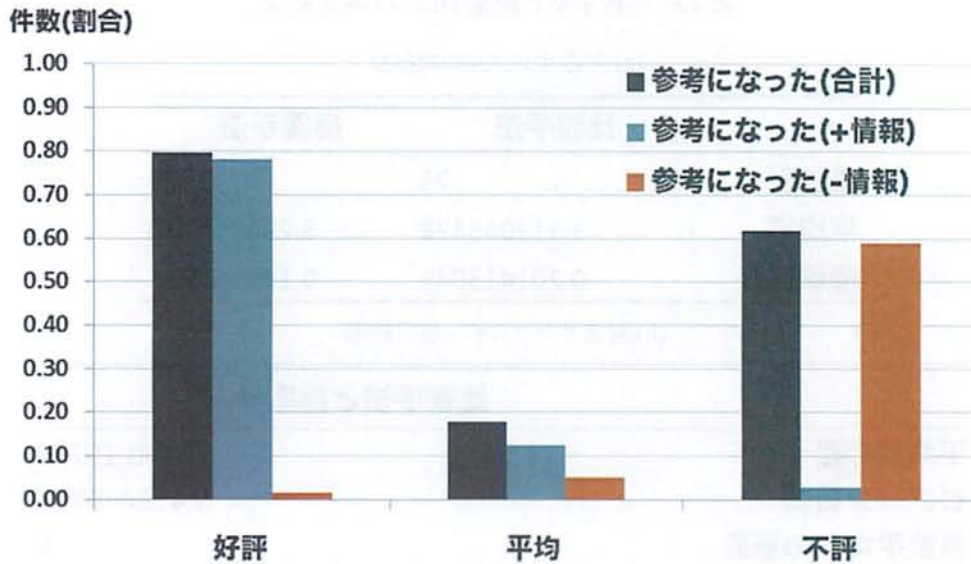


図 4.24: 属性の平均評価値に基づいた説明文において参考になった件数

い場合も暗に含むため、その属性を気に入らなかった場合にアイテムに満足しない可能性を懸念したためだと考える。同様に不評に基づいた説明文では「+情報」が全体の約5%しかないのに対し、 $lift_{ij}(neg \Rightarrow neg)$ では約21%を占めており、反対として役に立った意見の割合が多い。これも $lift_{ij}(neg \Rightarrow neg)$ が「この『属性』を気に入らない人は」という表現をしているため、その属性を気に入った場合にアイテムを満足する可能性を考えたからだと考察できる。このように説明文に条件文を使用することは、多様な解釈を生み出す可能性があるため、その表現には十分気をつける必要があると考える。

好評に基づいた説明文と $lift_{ij}(pos \Rightarrow pos)$ の説明文は、どちらも提示された説明文が示す属性を重要視する人にとっては、アイテムの利用価値を上げる意味で有効な提示情報になることが多い。その場合、前述したように好評に基づいた説明文の方が、一般的に簡潔でわかりやすい提示であると考えられる。一方、リフト値に基づく推薦文の場合、「多数の人が気に入っても自分が気に入るとは限らない」といった解釈も可能であるため、ユーザのこだわりに基づく価値判断に貢献していると考えられる。また、リフト値は3.2.2項で述べたように少数でも結びつきの強い組み合わせを検出するため、好評に基づいた説明文が作成しにくいような低評価なアイテムからでも、 $lift_{ij}(pos \Rightarrow pos)$ に基づいた説明文が作成可能な場合が多い。図4.26は、実際にそのような傾向がみられたアイテムの推薦説明の一部を示す。この例では好評として作成された説明文はなかったのに対し、 $lift_{ij}(pos \Rightarrow pos)$ は接客対応と食事に関して説明文が作成されていることがわかる。

またアイテムが低評価である場合、例えば「風呂」について「風呂が好評です」

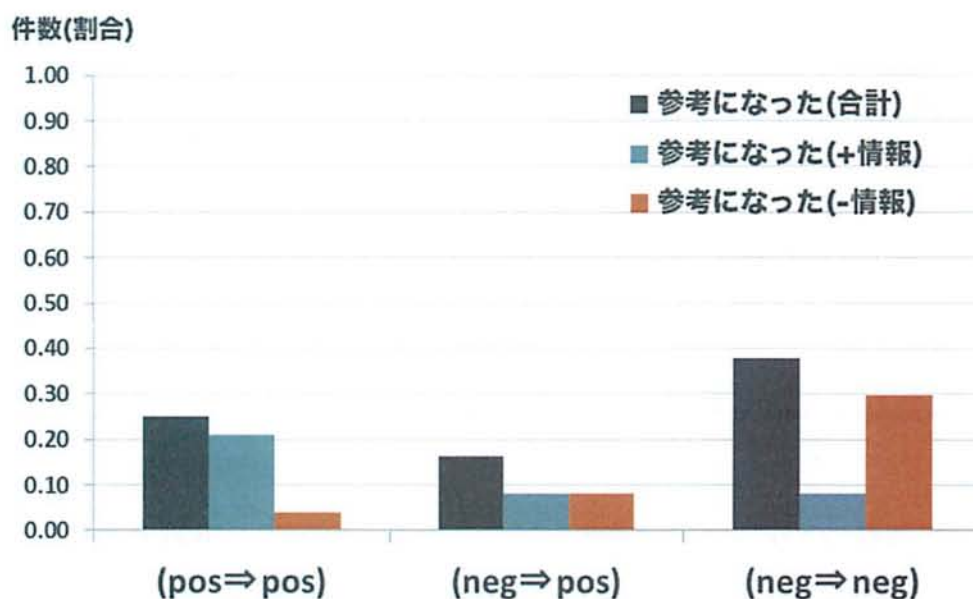


図 4.25: 価値観アイテムモデルに基づいた説明文において参考になった件数

と「風呂を好評とした人は、宿泊施設を気に入るでしょう」の説明文を比べると、風呂を重要視する人にとっては後者の表現の方が条件を強調した表現であり、より施設の利用価値を上げる可能性が期待できる。このことから低評価なアイテムでは、 $lift_{ij}(pos \Rightarrow pos)$ に基づいた推薦説明がより有効になる可能性があることが考察できる。同様に $lift_{ij}(neg \Rightarrow neg)$ に基づいた推薦説明も、図 4.27 のような高評価なアイテムにおいて有効になる可能性が考えられる。

図 4.28 に推薦アイテムと、実験で付与された評価値の分布を示す。図 4.28 の棒グラフは推薦アイテムに対し、レビューにより付与されている平均評価値（レビュー平均評価値）の分布をヒストグラムにして示したものであり、それらの各ビンに対して実際に協力者が実験で付与した評価値の平均（ユーザ平均評価値）を線グラフで示している。図 4.28 より提案手法は比較手法と比べて、推薦アイテムの持つ平均評価値の分布は広いことがわかる。その中でも 2.5 から 3 の相対的に低評価のアイテムも推薦しており、それらに対する協力者の平均評価値は約 3.5 となっていることがわかる。また図 4.29、図 4.30 に実際に提案手法で推薦されたアイテムとそれに対する協力者の回答の一部を示す。図 4.29、図 4.30 では、協力者が高い評価値を付与しているが、評価値だけでなくリフト値に基づく推薦説明が参考になったと回答している。このように提案手法では、従来の推薦手法では推薦対象とはならなかった、高評価ではないアイテムに対しても推薦できる可能性があり、ロングテールの推薦にも有効であることが期待できる。

宿泊施設の特徴

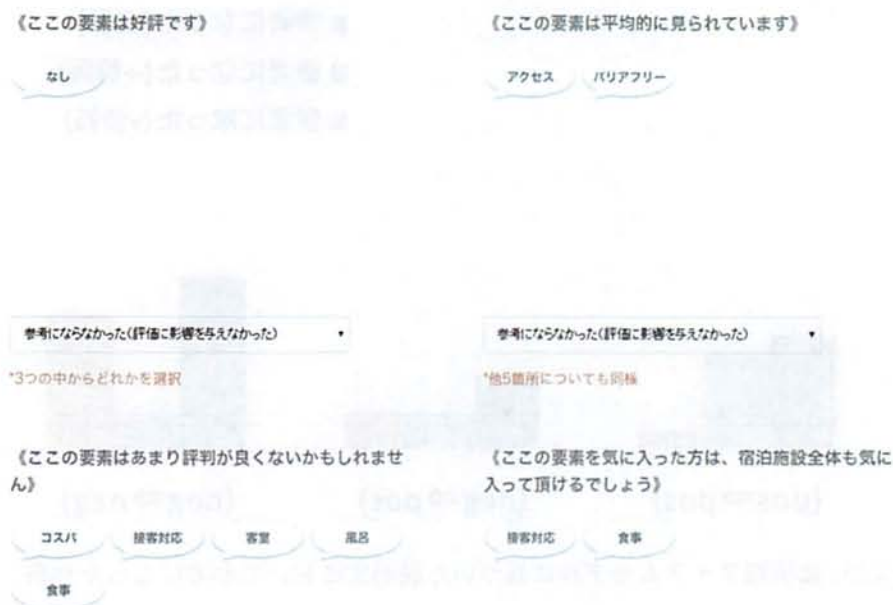


図 4.26: 低評価なアイテムにおける推薦説明の一部

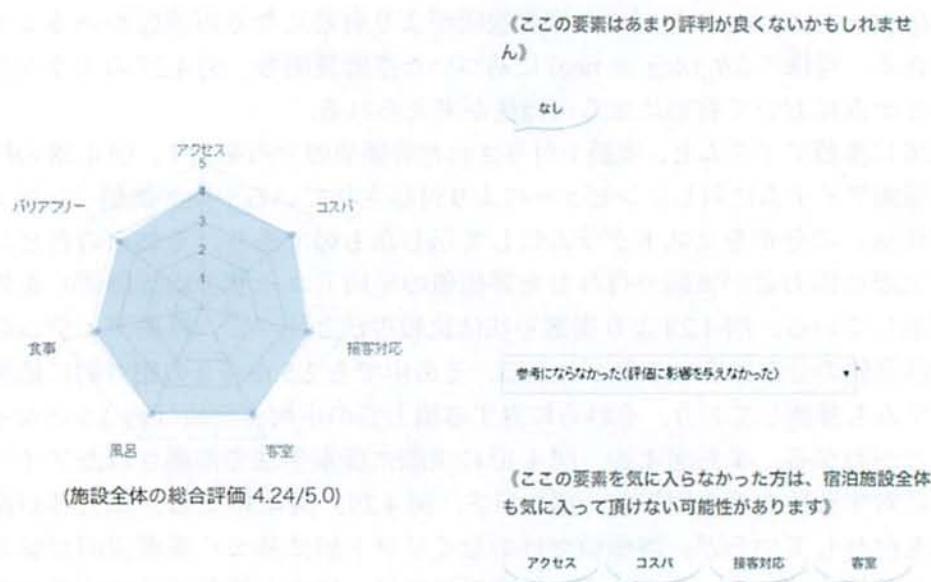


図 4.27: 高評価なアイテムにおける推薦説明の一部

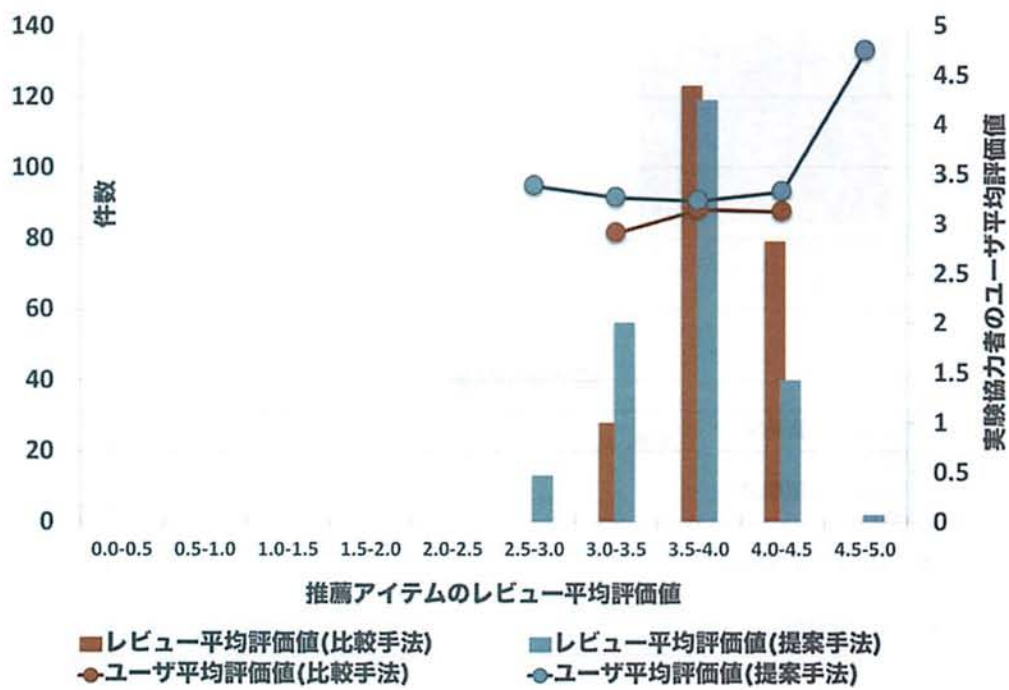


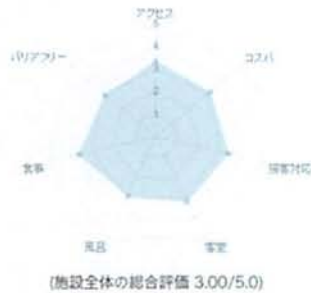
図 4.28: 推薦アイテムのレビュー平均評価値と付与されたユーザ平均評価値の分布図

伊勢志摩・二見 海辺のお宿 浜千代館 （旅館）

宿泊施設の基本情報



宿泊施設の簡単な説明：「和モダン」を基調としたレトロで静かな海辺のお宿。夫婦岩・二見シーパラダイスすぐ！伊勢神宮へも便利♪



協力者の回答

参考になった説明文 (+情報)	好評, $lift_{ij}(pos \rightarrow pos)$, $lift_{ij}(neg \rightarrow pos)$, $lift_{ij}(neg \rightarrow neg)$
参考になった説明文 (-情報)	なし
アイテムの平均評価値	3.0
付与した評価値	4.5

図 4.29: 提案手法による推薦アイテムと回答の一部, その 1

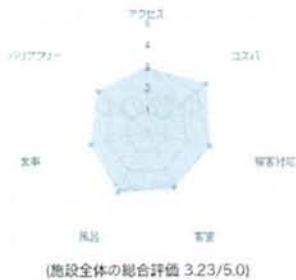
西伊豆松崎伊東園ホテル

[スタンダードホテル]

宿泊施設の基本情報



宿泊施設の簡単な説明：全室がオーシャンビューで、駿河湾を一望できる絶好のロケーション。



宿泊施設の特徴

《この要素は好評です》



《この要素を気に入った方は、宿泊施設全体も気に入って頂けるでしょう》



協力者の回答

参考になった説明文 (+情報)	好評, $lift_{ij}(pos \rightarrow pos)$
参考になった説明文 (-情報)	なし
アイテムの平均評価値	3.23
付与した評価値	4.0
評価値に対する理由	食事が高評価、かつ気に入れば全体も気に入るそうなので

図 4.30: 提案手法による推薦アイテムと回答の一部, その 2

5 おわりに

本論文では価値観に基づくアイテムモデリング手法、およびその推薦アルゴリズム、推薦説明への適用方法を提案した。提案手法の有効性を考察するため、提案手法を用いた推薦システムを構築し、評価実験によりその推薦精度や推薦説明に対する反応を収集した。

実験結果より、提案するアイテムモデルを用いた推薦手法は、従来のアイテムベースによる推薦と比べて推薦アイテムに対するユーザの平均評価値が大きいことがわかった。また相対的に好評でないがユーザが気に入るアイテムも推薦できしており、ロングテールの推薦にも有効である可能性を示した。価値観アイテムモデルに基づいた推薦説明の有効性についても考察し、好不評に基づく説明とは異なる性質を持つことや、少数のレビューからでも推薦の生成が可能であることを示した。

今後は提案したアイテム間の類似度計算方法を、予測評価値の計算以外にも利用することで、より多様なアイテムの推薦を行うことが実現可能と思われる。例えば特定のリフト値において類似度の高いアイテムを計算しておき、「～な見られ方をしているアイテム」とその類似アイテムの一覧を提示することで、アイテムの検索効率を向上させることが期待できる。また推薦説明においては、提案したアイテムモデルに、価値観に基づくユーザモデルを組み合わせることで推薦の理由文を作成することで、ユーザのシステムに対する理解度や満足度をさらに向上させることが期待できる。

謝辞

本研究は、著者が首都大学東京大学院システムデザイン研究科システムデザイン専攻情報通信システム学域において、多くの方々の御指導、ご協力の元に進めたものであります。

はじめに、指導教員である高間康史教授には、本研究の全般および学術論文にわたり、進行、執筆、発表に関する熱心な御指導、御助言を賜りました。本研究のみならず、重要事項の連絡など各方面において貴重な御指導、御協力を頂きました。ここに心より熱く御礼申し上げます。また本研究を行うにあたり多くの知識や助言を頂きました山口亨教授、小町守准教授、西谷隆夫教授にも深く感謝致します。

本研究の実験において御協力頂いた首都大学東京大学院の関係者に感謝致します。最後に、これまでの学生生活を理解し、暖かいご支援を頂いた家族、親友に心から感謝致します。

参考文献

- [1] B. M. Gross. The Managing of Organizations: The Administrative Struggle. No. 1 in The Managing of Organizations: The Administrative Struggle. Free Press of Glencoe, 1964.
 - [2] S. Hattori and Y. Takama. Recommender System Employing Personal-value-based User Model. Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol. 18, No. 2, pp. 157-165, 2014.
 - [3] 三澤 遼理, 服部 俊一, 高間 康史. 価値観に基づくユーザモデルの協調フィルタリングへの導入に関する検討. 第12回情報科学技術フォーラム (FIT2013), F-019, 2013.
 - [4] 三澤 遼理, 高間 康史. 価値観に基づくユーザモデルの協調フィルタリングへの適用に関する研究. 首都大学東京 卒業論文, 2014.
 - [5] R. J. Bayardo Jr and R. Agrawal. Mining the most interesting rules. 5th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 145-154, 1999.
 - [6] 神嶋 敏弘. 推薦システムのアルゴリズム (1). 人工知能学会誌, Vol.22, No.6, pp. 826-837, 2007.
 - [7] 神嶋 敏弘. 推薦システムのアルゴリズム (2). 人工知能学会誌, Vol.23, No.1, pp. 89-103, 2008.
 - [8] M. J. Pazzani and D. Billsus. Content-based recommendation systems. The Adaptive Web, Springer, pp. 325-341, 2007.
 - [9] 神嶋 敏弘. 推薦システムのアルゴリズム (3). 人工知能学会誌, Vol.23, No.2, pp. 248-263, 2008.
 - [10] 土方嘉徳. 嗜好抽出と情報推薦技術. 情報処理, Vol.48, No.9, pp. 957-965, 2007.
 - [11] G. Salton. Developments in automatic text retrieval. Science, Vol.253, pp. 974-979, 1991.
 - [12] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry. Using collaborative filtering to weave an information tapestry. Communications of the ACM, Vol. 35, No. 12, pp. 61-70, 1992.
-

- [13] J. S. Breese, D. Heckerman, and C. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. 14th conference on Uncertainty in artificial intelligence, pp. 43-52, 1998.
 - [14] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl. Grouplens: an open architecture for collaborative filtering of netnews. 1994 ACM conference on Computer supported cooperative work, pp. 175-186, 1994.
 - [15] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. 10th international conference on World Wide Web, pp. 285-295, 2001.
 - [16] G. Linden, B. Smith, and J. York. Amazon.com recommendations: Item-to-item collaborative filtering. IEEE Internet Computing, Vol. 7, No. 1, pp. 76-80, 2003.
 - [17] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. T. Riedl. An algorithmic framework for performing collaborative filtering. 22nd annual international ACM SIGIR conference on Research and development in information retrieval, pp. 230-237, 1999.
 - [18] M. Rokeach. The nature of human values. New York Free press, Vol. 438, 1973.
 - [19] M. B. Holbrook. Consumer value: a framework for analysis and research. Psychology Press, 1999.
 - [20] A. Roshchina, J. Cardiff, and P. Rosso. A comparative evaluation of personality estimation algorithms for the TWIN recommender system. 3rd International Workshop on Search and Mining User-generated Contents, pp. 11-18, 2011.
 - [21] R. Hu and P. Pu. A study on user perception of personality-based recommender systems. User Modeling, Adaptation, and Personalization, Springer, pp. 291-302, 2010.
 - [22] C. Schmitt, D. Dengler, and M. Bauer. Multivariate preference models and decision making with the MAUT machine. User Modeling 2003, Springer, pp. 297-302, 2003.
 - [23] J. S. Dyer, P. C. Fishburn, R. E. Steuer, J. Wallenius, and S. Zioints. Multiple criteria decision making, multiattribute utility theory: the next ten years. Management science, Vol. 38, No. 5, pp. 645-654, 1972.
-

- [24] A. Felfernig, K. Isak, K. Szabo, and P. Zachar. The vita financial services sales support environment. *National Conference on Artificial Intelligence*, Vol. 22, No. 2, p. 1692, 2007.
 - [25] M. A. S. Nunes and R. Hu. Personality-based recommender systems: An overview. In *Proceedings of the 6th ACM Conference on Recommender Systems*, pp. 5-6, 2012.
 - [26] J. M. Burger. *Personality (Psy 235 Theories of Personality)*. Cengage Learning, 2010.
 - [27] R. R. McCrae and P. T. Costa. A five-factor theory of personality. O. P. John, R. W. Robins, L. A. Pervin (eds.), *Handbook of personality: Theory and research*, Guilford Press, Vol. 2, pp. 139-153, 1999.
 - [28] J. M. Dusay. Egograms and the “constancy hypothesis.” *Transactional Analysis Journal*, Vol. 2, No. 3, pp. 37-41, 1972.
 - [29] R. R. McCrae, P. T. Costa Jr. F. Ostendorf, A. Angleiner, M. HÅebkov, M. D. Avia, and P. R. Saunders. Nature over nurture: temperament, personality, and life span development. *Journal of personality and social psychology*, Vol. 78, No. 1, p. 173, 2000.
 - [30] 清水涼人, 服部俊一, 高間康史. レビュー閲覧履歴からの価値観に関するユーザモデル構築手法の提案. 第28回人工知能学会全国大会 (JSAI2014), 3B4-OS-10b-3, 2014.
 - [31] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, Vol. 22, No. 1, pp. 5-53, 2004.
 - [32] S. M. McNee, J. Riedl, and J. A. Konstan. Being accurate is not enough: how accuracy metrics have hurt recommender systems. *CHI’06 extended abstracts on Human factors in computing systems*, pp. 1097-1101, 2006.
 - [33] C. N. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen. Improving recommendation lists through topic diversification. *14th international conference on World Wide Web*, pp. 22-32, 2005.
 - [34] L. Chen and P. Pu. Trust building in recommender agents. *Workshop on Web Personalization, Recommender Systems and Intelligent User Interfaces at 2nd International Conference on E-Business and Telecommunication Networks*, pp. 135-145, 2002.
-

- [35] N. Tintarev and J. Masthoff. A survey of explanations in recommender systems. 2007 IEEE 23rd International Conference on Data Engineering Workshop, pp. 801-810, 2007.
 - [36] S. W. Bennett, and A. C. Scott. Chap 19: Specialized Explanations for Dosage. B. G. Buchanan and E. H. Shortliffe (eds.), *The Rule-Based Expert Systems*, Addison-Wesley Publishing Company, pp. 363-370, 1985.
 - [37] H. Cramer, V. Evers, S. Ramlal, M. V. Someren, L. Rutledge, N. Stash, L. Aroyo, and B. Wielinga. The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction*, Vol. 18, No. 5, pp. 455-496, 2008.
 - [38] M. Czarkowski and J. Kay. A scrutable adaptive hypertext. *Adaptive Hypermedia and Adaptive Web-Based Systems*, Springer Berlin Heidelberg, pp. 384-387, 2002.
 - [39] D. Doyle, A. Tsymbal, and P. Cunningham. A review of explanation and explanation in case-based reasoning. Technical Report of Trinity College Dublin, Department of Computer Science, TCD-CS-2003-41, 2003.
 - [40] R. Ohanian. Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. *Journal of Advertising*, Vol. 19, No. 3, pp. 39-52, 1990.
 - [41] S. M. McNee, S. K. Lam, J. A. Konstan, and J. Riedl. Interfaces for Eliciting New User Preferences in Recommender Systems. Peter, C. Albert, and R. Fiorella (eds.), *User Modeling 2003*, Springer Berlin Heidelberg, pp. 178-187, 2003.
 - [42] J. L. Herlocker, J. A. Konstan, and J. Riedl. Explaining collaborative filtering recommendations. *ACM conference on Computer supported cooperative work*, pp. 241-250, 2000.
 - [43] D. Cosley, S. K. Lam, I. Albert, J. A. Konstan, and J. Riedl. Is seeing believing? :how recommender system interfaces affect users' opinions. *SIGCHI conference on Human factors in computing systems*, pp. 585-592, 2003.
 - [44] M. Bilgic and R. J. Mooney. Explaining recommendations: Satisfaction vs. promotion. *Beyond Personalization Workshop at, IUI*, Vol. 5, pp. 13-18, 2005.
 - [45] J. Vig, S. Sen, and J. Riedl. Tagsplanations: explaining recommendations using tags. *14th international conference on Intelligent user interfaces*, pp. 47-56, 2009.
-

- [46] C. A. Thompson, M. H. Goker, and P. Langley. A personalized system for conversational recommendations. *Journal of Artificial Intelligence Research*, Vol. 21, pp. 393-428, 2004.
 - [47] K. McCarthy, J. Reilly, L. McGinty, and B. Smyth. Thinking positively explanatory feedback for conversational recommender systems. *Explanation Workshop at European Conference on Case-Based Reasoning*, pp. 155-124, 2004.
 - [48] A. Felfernig and B. Gula. Consumer behavior in the interaction with knowledge-based recommender applications. *ECAI 2006 Workshop on Recommender Systems*, pp. 37-41, 2006.
 - [49] G. Carenini, V. O. Mittal, and J. D. Moore. Generating patient-specific interactive natural language explanations. *Annual Symposium on Computer Application in Medical Care*, pp. 5-9, 1994.
 - [50] N. Tintarev and J. Masthoff. Over and underestimation in different product domains. *ECAI Workshop on Recommender Systems*, pp. 14-19, 2008.
 - [51] C. Lewis and J. Rieman. The Task-Centered Design Process. C. Lewis and J. Rieman (eds.), *Task-centered user interface design: a practical introduction*, University of Colorado, pp. 7-12, 1994.
 - [52] R. Agrawal and R. Srikant. Fast Algorithms for Mining Association Rules. *20th VLDB Conference*, pp. 487-499, 1994.
 - [53] 服部 俊一. 価値観に基づくユーザモデルを用いた情報推薦システムに関する研究. 首都大学東京 博士論文, 2014.
 - [54] 山口 貴之, 服部 俊一, 高間 康史. 価値観アイテムモデリング手法を利用した推薦理由提示手法についての考察, 第9回 SIG-AM 研究会, 2015.
-

発表文献

- 山口 貴之, 服部 俊一, 高間 康史. 価値観アイテムモデリング手法を利用した推薦理由提示手法についての考察, 第9回 SIG-AM 研究会, 2015.
- T. Yamaguchi, S. Hattori and Y. Takama. Proposal of Personal-value-based Item Modeling and Its Application to Explanation of Recommendation. The 2015 Conference on Technologies and Applications of Artificial Intelligence, pp. 58-63, 2015.

概要

本論文では、価値観に基づくアイテムモデリング手法、及びその推薦アルゴリズム、推薦説明への適用方法を提案し、ユーザ実験によりその有用性を考察する。

近年、情報化技術の発展により、ユーザが膨大な情報の中から自分のニーズに合ったものを探すのが困難になるという問題が生じている。これに対する解決策の一つとして情報推薦システムが注目されている。情報推薦の代表的な手法に協調フィルタリングがあり、ユーザの行動履歴と類似する履歴を持つユーザの評価を利用して推薦アイテムを決定する。この手法はショッピングサイト等を含め幅広く使用されている。

情報推薦の手法の中にはユーザの嗜好だけでなく、意思決定の背後にある価値観に着目して推薦を行う手法が提案されており、ユーザベースの協調フィルタリングや内容ベースの推薦手法に対してその有効性が示されている。この価値観モデリング手法では、価値観をアイテムの各属性に対するユーザのこだわりとみなし、ユーザレビューよりユーザモデルを構築している。

本論文ではこの手法をアイテムモデリングに拡張することを目的とする。アイテムに対しては、アイテムに投稿されたレビューを収集し、そのアイテムがどの属性に着目して評価されているのかを価値観モデルとして求めることが可能と考える。価値観モデル構築に利用可能なレビューは、アイテムの方がユーザモデリングと比較して集めやすいことが期待できる。また、アイテム間の類似度計算に構築されたアイテムモデルを使用することでアイテムベースの推薦も可能となる。

さらに構築されたアイテムモデルをアイテムの特徴として捉えユーザに提示することで、推薦説明に活用する事も期待できる。情報推薦において、推薦アイテムと共にその推薦理由やアイテムの特徴を説明することは、推薦システムに対する信頼性や満足度を向上させる手段として、近年その必要性が高まってきており、価値観モデルに基づく推薦説明は、従来手法とは異なる観点からの情報提供に繋がることが期待できる。

本論文では価値観に基づくアイテムモデリング手法を提案する。また構築されたモデルを用いた、推薦アルゴリズム、推薦時のアイテム説明の生成手法も提案し、プロトタイプシステムを構築して有用性を検証する。アイテムモデルはモデル構築に多数のレビューを利用可能という利点から、相関ルールの指標の1つであるリフト値を算出し、アイテムの特性をより詳しく分析する。推薦アルゴリズムでは、アイテム間の類似度計算に提案するアイテムモデルを使用しアイテムベースの推薦を行う。さらに、構築されたアイテムモデルから説明文を作成、可視化することで推薦時にユーザに提示する。また事前にアイテムモデルの解釈に関する予備実験を行った結果に基づき、モデルを利用した推薦説明の作成に対する考察を行う。

本論文は5つの章から構成される。1章では、本論文における研究背景及び研究目的を記述する。2章では、関連研究として、代表的な推薦手法である内容ベースフィルタリングと協

調フィルタリングについて説明した後、価値観モデリングに関する研究、推薦説明に関する研究を紹介する。3章では、価値観に基づくアイテムモデリング手法を提案する。リフト値を用いたアイテムモデルの算出方法を説明した後、アイテムモデルを用いたアイテム間の類似度計算方法を述べる。さらにアイテムモデルを元にした推薦説明の作成方法も述べる。4章ではアイテムモデルの解釈に関する予備実験と、提案手法を実装した推薦システムによるユーザの評価実験の結果を示す。予備実験では、ユーザのアイテムモデルに対する解釈をアンケート形式で収集し、その結果を分析するとともに推薦システムへの適用方法を考察する。評価実験では推薦システムを利用して提案手法とベースラインにおける推薦精度を比較する。ベースラインにはピアソン相関を使ったアイテムベース協調フィルタリングを使用する。またアイテムの推薦時には提案手法による推薦説明の提示を行い、提案する推薦説明がユーザの意思決定に与える影響について考察する。5章では、提案手法についてまとめるとともに、今後の展望についても述べる。

Abstract

This thesis proposes a personal value based item modeling for recommender system. The proposed method extends existing personal value based user modeling, which models user's personal values using RMRate based on the polarities of evaluation. In order to analyze the property of items in more detail, the proposed method employs lift value instead of the RMRate. This thesis also proposes a method to generate explanation of recommendation, which is expected to improve user's satisfactions for recommender systems by showing process of recommendation.

A prototype recommender system is implemented, using which the effectiveness of the proposed method is evaluated by user experiments.