

NAIST-IS-DD0661004

博士論文

ユーザコンテキストを考慮した
情報推薦方式に関する研究

奥 健太

2009年3月31日

奈良先端科学技術大学院大学
情報科学研究科 情報生命科学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
博士(工学) 授与の要件として提出した学位論文である。

奥 健太

審査委員：

加藤 博一 教授 (主指導教員)
木戸出 正繼 教授 (副指導教員)
宮崎 純 准教授 (副指導教員)
中島 伸介 准教授 (京都産業大学)
土方 嘉徳 講師 (大阪大学)

ユーザコンテキストを考慮した 情報推薦方式に関する研究*

奥 健太

内容梗概

ユーザの嗜好およびユーザの状況（ユーザコンテキストとよぶ）に合った情報を提供する，コンテキスト依存型情報推薦方式を提案する．コンテキスト依存型情報推薦方式を実現するための技術的課題として以下の課題が挙げられる．

- (1) 膨大なアイテムの中からユーザの嗜好およびユーザコンテキストに合致したアイテムを選定する技術の確立．
- (2) 膨大な推薦候補アイテムに対し，ユーザの嗜好およびコンテキストに基づいた適切なランキングを提供する技術の確立．
- (3) 推薦時のコンテキストだけでなく，過去/未来のコンテキストを考慮したアイテムを選定する技術の確立．

本論文では，ユーザコンテキストに依存したユーザのアイテムに対する嗜好および価値判断基準を表現したモデルを提案することにより，上記の課題 (1) ~ (3) に取り組む．以下，本論文における各課題に対するアプローチについて述べる．

従来の情報推薦技術では，ユーザのアイテムに対する好みに関するデータを蓄積し，このデータに基づいてユーザ嗜好モデルを作成する．しかし，ユーザの嗜好はユーザコンテキストに応じて多様に変化するため，アイテムに対する好みに関するデータを入力とするだけでは適切なユーザ嗜好を表現することはできない．

* 奈良先端科学技術大学院大学 情報科学研究科 情報生命科学専攻 博士論文, NAIST-IS-DD0661004, 2009年3月31日.

そこで、課題 (1) に対しては、ユーザコンテキストに依存したユーザ嗜好を表現した、コンテキスト依存型ユーザ嗜好モデルを提案する。提案モデルを利用することにより、ユーザの嗜好およびユーザコンテキストに合致したアイテムを推薦候補として選定することが可能となる。

情報推薦技術の目的の一つとして、情報探索行為におけるユーザ負担の軽減がある。ユーザの情報探索行為における負担を軽減するために、ユーザへ提示する推薦アイテム数を少なくすることが重要である。そこで、課題 (2) に対しては、課題 (1) において選定された推薦候補アイテム集合に対するランキングを作成し、その上位 n 個のアイテムをユーザに提示する方法を取る。特に、ユーザが、おかれているコンテキストにおいて、アイテムのどの属性を基準に価値判断を行っているのかを表現するユーザのコンテキスト依存型価値判断基準モデルを提案する。提案モデルを利用することにより、推薦候補アイテム集合に対し、ユーザコンテキストに依存したユーザの価値判断基準に沿ったランキングをユーザに提供することが可能となる。

ユーザの嗜好は、ユーザがおかれているその時点での状況だけでなく、ユーザが過去に取った行動や将来予定している行動にも影響を受けると考えられる。故に、過去/未来のユーザの時系列的な行動履歴もユーザコンテキストの一部であると考えられる。したがって、コンテキスト依存型情報推薦において、過去/未来のユーザの時系列的な行動履歴を考慮に入れることは重要な課題である。課題 (3) に対しては、ユーザの行動履歴データから、推薦時のコンテキストに応じたユーザの行動パターンを表現した、コンテキスト依存型行動モデルを提案する。提案モデルを利用することにより、推薦時コンテキストに加え、過去/未来の時系列的なユーザの行動履歴も考慮した情報推薦が可能となる。

キーワード

情報推薦，コンテキスト依存，ユーザコンテキスト，ユーザ嗜好モデル

A recommendation method considering user's contexts*

Kenta Oku

Abstract

I propose a context-aware recommendation method which provides users with information suitable for their preference and situations that are contexts. The following are problems to construct the context-aware recommendation method:

- (1) To establish a filtering method selecting items suitable for users' preference and contexts from a mass of items.
- (2) To establish a ranking method providing ranking appropriate to users' contexts.
- (3) To establish a recommendation method providing items suitable for users' contexts including contexts in the past / future.

In this thesis, I propose models representing users' preference and sense of values for items depending on the users' contexts in order to tackle the above problems. I show my approaches to the problems in this thesis.

Conventional recommendation methods make a users' preference model based on data on the users' preference of items. However, it is inadequate to consider only the preference of items, because the users' preference is changed according to their contexts. Thus, I propose a context-aware users' preference model

* Doctoral Dissertation, Department of Bioinformatics and Genomics, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DD0661004, March 31, 2009.

which represents the users' preference depending on their contexts for the problem (1). My proposed method can select items suitable for the users' preference and contexts as recommendation candidates by utilizing this model.

One of the purposes of the recommendation method is to lighten users' burden in information search. For that reason, it is important to reduce items to be presented to the users in order to lighten the users' burden. Therefore, I employ a way to present the top n items to the users based on ranking for recommendation candidates selected in the problem (1). Particularly, I propose a context-aware users' values model which represents the users' sense of values in item selection depending on their contexts for the problem (2). My proposed method can provide ranking for recommendation candidates based on the users' sense of values depending on their contexts by utilizing this model.

Users' preference is also changed according to the users' actions that have taken / will take in the past / future besides the users' current contexts. Hence, the users' action history in the past / future is also one of the users' contexts. Therefore, it is important to take the users' action history in the past / future into account to construct the context-aware recommendation method. I propose a context-aware users' actions model which represents the users' action pattern depending on their current contexts from their action history data for the problem (3). My proposed method can recommend items considering the users' action history in the past / future besides their current contexts by utilizing this model.

Keywords:

recommendation, context-awareness, users' contexts, users' preference model

目次

1. 序論	1
2. 関連研究	7
2.1 情報推薦技術の関連研究	7
2.1.1 コンテンツベースフィルタリング方式	7
2.1.2 協調フィルタリング方式	8
2.2 コンテキストウェアシステムの関連研究	10
3. ユーザコンテキストの定義とコンテキスト依存型情報推薦システム	13
3.1 緒言	13
3.2 ユーザコンテキストの定義	13
3.3 コンテキスト依存型情報推薦システム	15
4. コンテキスト依存型ユーザ嗜好モデルに基づく情報推薦方式	17
4.1 緒言	17
4.2 基礎的事項	18
4.3 コンテキスト依存型ユーザ嗜好モデル化方式	19
4.3.1 ユーザ嗜好のモデル化手順	20
4.3.2 ユーザ嗜好モデルに基づくユーザ類似度の算出手順	22
4.4 コンテキスト依存型ユーザ嗜好モデルに基づく情報推薦方式	24
4.4.1 コンテキスト依存型情報フィルタリング方式 (C-IF)	24
4.4.2 コンテキスト依存型協調フィルタリング方式 (C-CF)	26
4.5 ユーザ嗜好モデル化方式としてSVMを適用することの妥当性の検証	27
4.5.1 実験方法	28
4.5.2 実験結果と考察	29
4.6 被験者満足度による推薦方式の評価	30
4.6.1 実験条件	31
4.6.2 ランダム提示およびコンテキスト非依存の推薦方式との比較	33
4.6.3 C-IF および C-CF の推薦対象エリア別の有効性の比較	36

4.6.4	推薦対象エリア特性に基づく C-IF の有効性の分析	38
4.7	特徴パラメタの最適化に関する検討	39
4.7.1	実験方法	39
4.7.2	実験結果	40
4.7.3	考察	41
4.8	結言	43
5.	コンテキスト依存型ランキング方式	45
5.1	緒言	45
5.2	基本的事項	46
5.3	コンテキスト依存型ランキング方式	47
5.3.1	ユーザのコンテキスト依存型価値判断基準モデル	48
5.3.2	ユーザの学習データからの価値判断基準モデルの構築	50
5.3.3	コンテキスト依存型価値判断基準モデルに基づくランキン グ方式	62
5.4	評価実験	63
5.4.1	実験条件	63
5.4.2	実験方法	64
5.4.3	各ランキング方式との比較	79
5.4.4	ユーザの価値判断基準モデル構築の対象とするアイテムパ ラメタ選定に関する考察	82
5.4.5	ユーザの価値判断基準に影響を及ぼすコンテキストパラメ タ選定に関する考察	84
5.5	結言	85
6.	ユーザコンテキストの時系列性を考慮した情報推薦方式	87
6.1	緒言	87
6.2	基礎的事項	88
6.2.1	行動パターン抽出手法	88
6.2.2	ユーザの行動パターンに沿った情報推薦技術	90

6.3	提案方式の基本アイデア	92
6.3.1	概要	92
6.3.2	問題設定	92
6.3.3	推薦手順	96
6.4	行動パターン抽出手法の適用	107
6.4.1	シーケンシャルパターンマイニング手法の適用例	107
6.4.2	LCS 抽出手法の適用例	109
6.5	評価実験	111
6.5.1	実験条件	111
6.5.2	実験方法	112
6.5.3	実験結果および考察	119
6.6	結言	125
7.	結論	127
7.1	論文のまとめ	127
7.2	コンテキスト依存型情報推薦システムの適用例	129
7.3	コンテキスト依存型情報推薦システムの運用指針	130
7.4	ユーザのデモグラフィック情報の利用	132
7.5	今後の展望	133
	謝辞	135
	参考文献	139

目 次

1	本論文で提案する情報推薦システムの構成図	16
2	SVM の特徴空間	19
3	コンテキスト依存型ユーザ嗜好モデル化方式	20
4	CP モデルに基づくユーザ類似度の算出	23
5	コンテキスト依存型情報フィルタリング方式 (C-IF) による推薦手順	25
6	コンテキスト依存型協調フィルタリング方式 (C-CF) による推薦 手順	27
7	実データセットを対象としたときの未知データに対する識別性能 の比較	31
8	人工データセットを対象としたときの未知データに対する識別性 能の比較	31
9	各推薦方式に対する被験者満足度の比較	35
10	既知/未知エリアにおける C-IF および C-CF に対する満足度の比較	37
11	各被験者グループの推薦対象エリア別の C-IF に対する満足度の比較	38
12	特徴パラメタがモデルの識別性能に与える影響の分析	41
13	ユーザのアイテム選択に対する価値判断基準 (飲食店選択の例)	49
14	ユーザのアイテム利用履歴データと提案方式で扱う学習データ形 式への変換	51
15	周期型パラメタの表現 ([時刻] の例)	53
16	アイテムパラメタ軸上における正負データの分布に基づく価値判 断基準モデルの構築	57
17	アイテムパラメタの重要度の算出例とデータペアの重なり判定 (連 続値型パラメタの場合)	58
18	アイテムパラメタの重要度の算出例 (2 値型およびカテゴリ型の場 合)	60
19	正負データの分布に基づく価値判断基準モデルの構築手順	61
20	推薦候補アイテムに対するランキングスコアの算出	62
21	実験用コンテキストパターンの設定例 (被験者 A の場合)	66

22	実験で用いた学習データセット (被験者 A の場合)	67
23	実験で用いた評価用データセット (被験者 A の場合)	69
24	予備実験 : 学習データ数と未知データに対する識別性能の関係 . . .	72
25	アイテムパラメタ x_{13} 軸上における正負学習データの分布 (コンテ キスト C_1 における被験者 A の場合)	73
26	アイテムパラメタ x_{13} に対するユーザの価値判断基準モデル (コン テキスト C_1 における被験者 A の場合)	75
27	提案方式による評価用データセットに対するランキング (被験者 A の場合)	76
28	ランキング手法の比較	80
29	各方式による評価用データセットに対するランキング (コンテキス ト C_1 における被験者 A の場合)	81
30	アイテムパラメタ選定によるランキング結果に対する DCG の値の 違い	83
31	コンテキストパラメタ選定によるランキング結果に対する DCG の 値の違い	85
32	ユーザ $\{B, C, D, E\}$ の行動履歴データベース	98
33	ユーザ $\{B, C, D, E\}$ のユーザ A とのユーザ類似度	100
34	行動の予測スコアの算出例 (行動 “バーで飲む” の場合)	106
35	被験者 A による実験の例	115
36	本実験における被験者間のユーザ類似度	117
37	推薦時コンテキスト考慮時/非考慮時における SPM によるランキ ング結果に対する R 適合率	119
38	推薦時コンテキスト考慮時/非考慮時における LCS によるランキ ング結果に対する R 適合率	120
39	類似ユーザを考慮したときの SPM によるランキング結果に対する R 適合率	121
40	類似ユーザを考慮したときの LCS によるランキング結果に対する R 適合率	122

41	SPM における最小サポート値による R 適合率	123
42	ネガティブ行動パターン考慮/非考慮時の R 適合率	124

表 目 次

1	ユーザコンテキストの例 (繁華街における回遊情報を推薦対象コン テンツとした場合)	14
2	クラス分類手法の性能分析に用いる人工データセット	29
3	本実験で用いた飲食店データセット	32
4	本実験で用いた飲食店特徴パラメタ (28 次元)	33
5	本実験で用いたコンテキスト特徴パラメタ (24 次元)	33
6	特徴パラメタを累積的に排除したときの識別性能の変化	42
7	実験で用いた学習データ形式	64
8	各コンテキストパラメタに対する情報利得 (コンテキスト C_1 にお ける被験者 A の場合)	71
9	各アイテムパラメタの重要度 (コンテキスト C_1 における被験者 A の場合)	74
10	ユーザの行動履歴データベース (例)	95
11	ユーザの行動シーケンスデータベース (例)	95
12	推薦時コンテキスト C^A に関連した行動履歴データベース	102
13	推薦時コンテキスト C^A に関連した行動シーケンスデータベース	102
14	各行動シーケンスの出現頻度と行動パターン	103
15	過去コンテキスト T^A に合った頻出行動パターン	105
16	各行動の予測スコア	106
17	シーケンシャルパターンマイニング手法による行動パターンの抽 出結果 (minsup = 0.2)	108
18	シーケンシャルパターンマイニング手法による過去コンテキスト に合った行動パターンの特定結果	109
19	シーケンシャルパターンマイニング手法による各行動の予測スコア	109

20	LCS 抽出手法による行動パターンの抽出結果 ($min.Count = 1, min.Length = 2$)	110
21	LCS 抽出手法による過去コンテキストに合った行動パターンの特定結果	110
22	LCS 抽出手法による各行動の予測スコア	111
23	実験で用いた施設データ	112
24	実験で用いた推薦時コンテキストおよび過去コンテキスト	113
25	各被験者の推薦時コンテキストの設定	114
26	被験者 A の評価データに対する行動予測スコア (推薦時コンテキスト=「平日の 20 時に一人で」, 過去コンテキスト=「(19 時, ”食事する”)」の場合)	118

1. 序論

近年，Web の普及に伴い，ユーザが取得できる情報量が大幅に増加している．一方で，Web 上にはユーザにとって不要な情報も大量に散在し，膨大な情報の中から適切な情報を効率的に検索することが困難になっている．したがって，ユーザにとって適切な情報を効率的に検索する手法を開発することに対する要求は高い．

このような背景の中，情報推薦技術に関する研究が注目されている．情報推薦技術は，膨大な情報の中からユーザの嗜好に合った情報を見つけ出し，それをユーザに提供する技術である．この情報推薦技術は，Amazon.com[1] や Google News[2] など，実際に推薦サービスとして実現され，多くのユーザに日常的に利用されている．

一方で，近年では，携帯電話をはじめとする，ノート型パソコンや PDA (Personal Digital Assistant)，携帯型デジタル音楽プレイヤー等，モバイル機器の普及に伴い，個人が情報機器に触れる場面も多種多様なものとなってきた．このため，ユーザはいつでも，どこでも，情報サービスを利用できるようになった．このような中，ユーザ側の要求としては，そのときの状況，すなわちユーザコンテキストに合った情報を提供して欲しいというものがある．例えば，「ドライブ中に，雰囲気合った音楽を再生してほしい」，「一人で街に出かけているときに，できるだけ近くの安いレストランを探したい」，「観光地で，次に巡回すべきお勧めの観光施設の情報を知りたい」などといった場面が思い浮かばれる．したがって，ユーザコンテキストに応じた情報推薦技術を研究・開発することに対する意義は大きい．

そこで，本論文では，「情報選択過程におけるユーザの嗜好に影響を及ぼす状況」をユーザコンテキストと定義したうえで，ユーザの嗜好だけでなくそのときのユーザコンテキストに合った情報を提供する，コンテキスト依存型情報推薦システムの実現を目標とする．コンテキスト依存型情報推薦システムを実現するための技術的課題として，以下のものが挙げられる．

- (1) ユーザコンテキストに依存したユーザ嗜好のモデル化方式の確立．
- (2) 推薦候補アイテム集合に対し，ユーザコンテキストに依存したランキング

方式の確立．

- (3) 過去/未来といったユーザコンテキストの時系列性を考慮した情報推薦方式の確立．

以下，各課題と各課題に対する本論文のアプローチについて述べる．

(1) ユーザコンテキストに依存したユーザ嗜好のモデル化方式の確立

課題

従来の情報推薦技術を実現する方式として，最も主流なものの一つとして，ユーザ嗜好モデルを用いた方法がある [3][4]．ユーザ嗜好モデルは，基本的には，特徴化されたアイテムデータとアイテムに対する嗜好データ（好き/嫌いなど）を学習データとした，機械学習手法により作成される．

しかしながら，ユーザの嗜好はユーザコンテキストによって大きく変化するものであり，これを適切に学習し，ユーザ嗜好モデルとして表現することは容易ではない．ユーザコンテキストに応じて変化するユーザの嗜好をどのように学習するかということが大きな課題である．

最も単純な解としては，考えられるコンテキスト別に学習データを用意し，これに基づいたユーザ嗜好モデルをコンテキスト別に作成することで，各コンテキストに応じた推薦を行う方法が考えられる．しかしながら，想定されるコンテキストのパターン数は，膨大な数に上り，各コンテキストに逐一对応した学習データを用意することは現実的でない．また，学習時に想定していないコンテキストにおいては，適切な情報推薦を行うことはできない．

したがって，あらゆるユーザコンテキストに対し，柔軟に対応した，ユーザ嗜好モデルを適切に作成する方式を確立することが課題となる．

本論文のアプローチ

本論文では，ユーザコンテキストに応じて変化するユーザの嗜好に対応した，コンテキスト依存型ユーザ嗜好モデルを提案する．

提案モデルでは，アイテムデータに加え，コンテキストデータも特徴化する．そして，特徴化されたアイテムデータとコンテキストデータ，おかれたコンテキストにおけるアイテムに対する嗜好データを学習データとした，機械学習手法によりユーザ嗜好モデルを作成する．こうすることにより，例えば，あるコンテキスト s における対象アイテム A (A_s と表現) と，コンテキスト t における対象アイテム A (A_t と表現) を別のデータとして扱うことができる．アイテムデータおよびコンテキストデータを併せた学習データを用いることにより，コンテキストに依存したユーザの嗜好を学習することが可能となる．さらに，あらかじめコンテキストデータも特徴化しておくことにより，未知のコンテキストにおいても柔軟にユーザ嗜好の推測を行うことができる．

本論文では，コンテキスト依存型ユーザ嗜好モデルの提案とともに，このモデルに基づく以下の二つの情報推薦方式を提案する．

- (a) コンテキスト依存型情報フィルタリング方式 (Context-Aware Information Filtering Method; C-IF)
- (b) コンテキスト依存型協調フィルタリング方式 (Context-Aware Collaborative Filtering Method; C-CF)

これらの情報推薦方式を用いることにより，ユーザの嗜好およびコンテキストに適切なアイテムを推薦候補として選定することが可能となる．

また，飲食店情報推薦を想定した被験者実験により，提案のモデルおよび情報推薦方式の有効性を検証する．

(2) 推薦候補アイテム集合に対し，ユーザコンテキストに依存したランキング方式の確立．

課題

情報推薦技術の目的の一つとして，情報探索行為におけるユーザ負担の軽減がある．ユーザの情報探索行為に対する負担を軽減するために，課題 (1) において選定された推薦候補アイテム集合に対するランキングを作成し，ユーザに提示するアイテム数を上位いくつかに絞り込む．

ランキングを提供する従来方式としては，ベクトル空間モデル [5][6][7] を用いた方法や多クラス分類器 [8][9] を用いた方法などが挙げられる．

ベクトル空間モデルでは，あらかじめユーザ嗜好と推薦候補となるアイテムを特徴ベクトルとして表現しておく．そして，ユーザ嗜好の特徴ベクトルとの類似度が大きいアイテムを上位とすることで，ランキングを提供する．しかし，課題 (1) においても述べたように，そもそもユーザコンテキストに依存するユーザ嗜好を適切に表現できていなければ，ランキング結果に対する信頼性も乏しい．

また，多クラス分類器では，あらかじめユーザ嗜好の程度に基づいて順序付けられたアイテムを学習データとすることにより，アイテム間の順序関係を含めたモデルを作成する．しかし，ユーザのアイテムに対する嗜好はユーザコンテキストに応じて変化するため，ユーザコンテキストが変わればアイテム間の順序関係も変わってしまう．したがって，事前にアイテム間の順序関係を定義することは不可能である．

したがって，ユーザの嗜好がユーザコンテキストによって変化することを考慮に入れた，ランキング方式を確立することが課題となる．

本論文のアプローチ

本論文では，ユーザの価値判断がユーザコンテキストによって変化することを考慮に入れて，ランキングを行う，コンテキスト依存型ランキング方式を提案する．

コンテキスト依存型ランキング方式の流れは，次のとおりである．

- (1) ユーザの学習データからのコンテキスト依存型価値判断基準モデルを構築する．
- (2) 価値判断基準モデルに基づき，推薦候補アイテム集合をランキングする．

まず，ユーザの学習データから，コンテキスト依存型価値判断基準モデルというものを作成する．このモデルは，推薦時のユーザコンテキストにおいて，ユーザがアイテムのどの属性を重視して，アイテムに対する価値判断を行っているかを表現するものである．このモデルに基づいて，推薦候補アイテム集合に対する

スコア付けを行うことにより，ユーザコンテキストに合ったランキングを提供することが可能となる．

また，本論文では，飲食店情報推薦を想定した被験者実験により，提案のランキング方式の有効性を検証する．

(3) 過去/未来といったユーザコンテキストの時系列性を考慮した情報推薦方式の確立．

課題

ユーザの嗜好は，ユーザがおかれているその時点での状況だけでなく，ユーザが過去に取った行動や将来予定している行動にも影響を受けると考えられる．故に，過去/未来のユーザの時系列的な行動履歴もユーザコンテキストの一部であると考えられる．したがって，コンテキスト依存型情報推薦において，ユーザコンテキストの時系列性を考慮に入れることは重要な課題である．

従来研究として，ユーザの行動パターンに沿った情報推薦 [10][11] が挙げられる．これら従来研究では，基本的には，蓄積されたユーザの行動履歴データから行動パターンを抽出し，直近のユーザの行動と類似する行動パターンに沿って推薦情報を決定している．人には日常的に何らかの行動パターンがあり，この行動パターンを考慮することで，ユーザにとって有益な情報推薦が可能となる．

ただし，ユーザの行動パターンも推薦時のユーザコンテキストに応じて変化するため，ユーザの一般的な行動パターンを抽出するだけでは不十分である．ユーザコンテキストに依存した行動パターンを抽出することが重要となる．

本論文のアプローチ

本論文では，過去/未来のユーザの行動もユーザコンテキストとして扱った，コンテキスト依存型情報推薦方式を提案する．

提案する情報推薦方式では，まず，ユーザの行動履歴データベースから，推薦時コンテキストに応じた限定的な行動パターンの抽出を行う．そして，この限定的な行動パターンに基づき，推薦情報の判定を行う．推薦時コンテキストに応じた限定的な行動パターンを考慮することにより，推薦時コンテキストおよび過去/

未来のコンテキストの両方を考慮した情報推薦を行うことが可能となる。

また、繁華街における回遊行動を想定した被験者実験により、提案の情報推薦方式の有効性を検証する。

本論文は、以下の7章から構成される。

第2章では、情報推薦技術およびコンテキストウェアシステムの従来に関連研究を整理し、本研究の位置付けを述べる。

第3章では、本論文におけるユーザコンテキストの定義を明確にする。また、本論文で提案するコンテキスト依存型情報推薦システムの全体像を示し、第4章および第5章、第6章の位置付けを明確する。

第4章では、上記課題(1)への取組みとして、コンテキスト依存型ユーザ嗜好モデルおよびこのモデルに基づいた情報推薦方式を提案する。そして、被験者実験を通じて、提案方式の有効性について検証および考察を行う。

第5章では、上記課題(2)への取組みとして、コンテキスト依存型価値判断基準モデルおよびこのモデルに基づいたランキング方式を提案する。そして、被験者実験を通じて、提案方式の有効性について検証および考察を行う。

第6章では、上記課題(3)への取組みとして、ユーザコンテキストの時系列性を考慮した情報推薦方式を提案する。そして、被験者実験を通じて、提案方式の有効性について検証および考察を行う。

最後に、第7章で、本論文で得られた成果をまとめる。

2. 関連研究

2.1 情報推薦技術の関連研究

情報推薦技術に関する研究は，1990年代半ばに Resnick ら [12] によって，協調フィルタリングの概念が初めて定式化されて以来，盛んに行われるようになった [3][4] .

情報推薦技術では，基本的にはユーザの過去の行動履歴からユーザの嗜好をモデル化し，そのモデルに基づいた推薦を行う．情報推薦の基本方式には，大きく以下の2種類がある [13][14] .

- (a) コンテンツベースフィルタリング方式
- (b) 協調フィルタリング方式

以下，各方式について説明する．

2.1.1 コンテンツベースフィルタリング方式

コンテンツベースフィルタリング方式では，ユーザが過去に好んだアイテムに類似するアイテムがユーザに推薦される [15][16][17] . 特に，推薦候補となるアイテムと過去にユーザが評価付けしたアイテムとを比較することにより，ユーザにとって最適なアイテムが推薦される．例えば，映画コンテンツを推薦対象とした場合，ユーザが過去に評価付けした映画に関する情報から，ユーザの映画に対する嗜好 (好きな俳優，好きな監督，好きなジャンルなど) を抽出する．そして，その嗜好に合った映画をそのユーザに推薦する．

コンテンツベースフィルタリング方式は，大きく次の2種類に分類される．

- メモリベース方式
- モデルベース方式

メモリベース方式では，ユーザプロフィールに基づいた推薦を行う．ユーザプロフィールとは，ユーザの興味に関する情報を記述したものであり，通常，ユーザ

が過去に閲覧および評価したアイテムの内容に基づいて作成される。ユーザプロフィールの一つの表現方法として、ベクトル空間モデル [5][6][7] を用いる方法がある。例えば、ユーザプロフィールは重みベクトル $(w_{c1}, w_{c2}, \dots, w_{cl})$ により表現される。ここで、 w_{ci} は、ユーザ c のキーワード k_i に対する重みを表す。Rocchio アルゴリズム [6] では、ユーザが過去に好んだアイテムの特徴ベクトルの平均ベクトルとして、ユーザプロフィールを作成している [13][15]。

モデルベース方式は、ユーザが過去に利用したアイテムに対する満足度評価からユーザ嗜好モデルを作成する。ここで、アイテムの特徴は多次元の特徴ベクトル形式などにより表現される。このとき、ユーザ嗜好モデルとアイテムの特徴とを比較することにより、ユーザが好みそうなアイテムを推測する。一般的には、Support Vector Machine[18] やベイズ分類器 [19][20]、クラスタリング手法 [21]、決定木 [22][23]、ニューラルネットワーク [24] などといった、機械学習手法を用いてユーザ嗜好モデルが作成される。

Pazzani ら [17][25] や Littlestone ら [26] は、文書に対するユーザの嗜好をベイズ分類器によりモデル化している。岩濱ら [27] は、決定木を用いて音楽コンテンツに対するユーザの嗜好をモデル化している。

以上のように、コンテンツベースフィルタリング方式では、基本的にはユーザの過去のアイテムに対する好みに関するデータに基づいて、ユーザプロフィールやユーザ嗜好モデルが作成され、これらに合致したアイテムが推薦候補として選定される。しかし、本論文ではユーザ嗜好がユーザコンテキストに応じて多様に変化するという点に着眼しており、アイテムに対する好みに関するデータからだけでは、一概にどのようなコンテキストにも対応可能なユーザ嗜好モデルを作成することはできないと考える。したがって、コンテキスト依存型情報推薦方式を確立するためには、ユーザコンテキストをも含めたユーザ嗜好モデルの作成が必要となる。

2.1.2 協調フィルタリング方式

協調フィルタリング方式では、アイテムに対するユーザの嗜好を他のユーザの嗜好データに基づいて予測する。例えば、映画コンテンツを推薦対象とした場合、

まず，対象ユーザと映画に対する嗜好が類似する他のユーザ（類似ユーザ）を見つけ出す．そして，その類似ユーザが好む映画を対象ユーザに推薦する．

協調フィルタリング方式を用いた推薦システムは，これまでも多くのものが開発されている．GroupLens[12][28] や Video Recommender[29]，Ringo[30] は，協調フィルタリング方式を用いた初の情報推薦システムである．実用サービスとして提供されているものとして，Amazon.com[1][31] や Google News[2]，TSUTAYA online[32] などがある．

協調フィルタリング方式は，大きく次の 2 種類に分類される [33]．

- メモリベース方式
- モデルベース方式

メモリベース方式 [33][34] [35][12][30] では，あらかじめユーザが過去に与えたアイテムに対する評価値に基づいて，ユーザ間の類似度が求められる．この類似度が最も高い上位 N 人のユーザのデータに基づき，未評価であるアイテムの評価値予測を行う．そして，予測された評価値が高いアイテムをユーザに推薦する．

モデルベース方式 [36][33] [37][38][39] [40][41] では，クラスタリング手法などにより，ユーザ間の関係をモデル化しておく．つまり，同一クラスタに属するユーザは，互いに嗜好が類似することになる．このクラスタに基づいて，ユーザへの推薦を行う．

以上のように，協調フィルタリング方式では，ユーザ類似度の算出方法の定義が重要となる．基本的には，ユーザのアイテムに対する評価値に基づいて，ユーザ類似度が算出される．しかし，先に述べたことと同様，ユーザの嗜好はユーザコンテキストに依存するため，ユーザのアイテムに対する評価値はユーザコンテキストによって変わる．このため，アイテムに対する評価値だけからでは，正確な嗜好の類似度は求められないと考えられる．したがって，コンテキスト依存型情報推薦方式を確立するためには，評価時のユーザコンテキストを含めたユーザ類似度の算出が必要となる．

2.2 コンテキストウェアシステムの関連研究

ノートやパソコンやPDA (Personal Digital Assistant), 携帯電話, 携帯型デジタル音楽プレイヤーなどのようなモバイル機器の普及に伴い, コンテキストウェアシステムが注目されるようになった [42]. コンテキストウェアシステムは, ユーザからの明示的な入力がなくとも, そのときのコンテキスト (時刻や場所, 環境など) に応じたサービスをユーザに提供するものである. コンテキスト情報は, センサやネットワーク, モバイル機器の状態など, さまざまな方法により取得される.

コンテキストウェアシステムに関する研究は, Wantら [43] が Active Badge Location System を提案したことから始まった. 1990年代後期には, ユーザの位置に合わせた情報を提供する位置依存型の旅行ガイドシステム (Abowdら [44], Sumiら [45], Cheverstら [46]) が現れた. また, 森下ら [47] は, 時空間限定でアクセス可能な仮想オブジェクトである SpaceTag を提案している. これは, ユーザが特定の時間に特定の場所に訪れると, そのオブジェクトにアクセスすることができるというものである. ユーザの位置情報に関しては, GPS などにより容易に取得できることから, コンテキストウェアシステムの一つとして位置依存型のシステムが盛んに研究されてきた.

2000年代初期には, コンテキスト情報に加え, ユーザの興味などを表すユーザプロフィールも考慮に入れた, Context-Aware Recommendation に関する研究が行われるようになった. モバイルユーザを対象とした推薦システムである, PILGRIM[48] や MALCR[49] では, ユーザの位置情報やユーザプロフィールに合った Web ページや広告をユーザに提供する. COMPASS[50] では, 現在位置から一定の範囲内にあるレストランの中から, ユーザがあらかじめ指定した興味ジャンルに合致するレストランをユーザに提示する.

2000年代中期からは, ユーザの位置情報や時間情報だけでなく, 他の情報もコンテキストとして扱った推薦システムも研究されるようになってきた. Kwon[51] は, ユーザが閲覧中の Web ページ内にある画像をコンテキストとして扱い, その画像の内容に関連付けられた広告をユーザに提示する推薦システム, CAMA を提案している. Yuら [52] は, ユーザの持つモバイル端末の性能やネットワークの

帯域幅をコンテキストとして扱い、それに合わせて、コンテンツのユーザへの提示形式(テキスト形式, 静止画形式, 動画形式)を選択する推薦システム CoMeR を提案している。

このように、コンテキストウェアシステムに関する研究が進むにつれて、扱うコンテキストの種類も増加してきた。このような中、コンテキストウェアシステムを確立させるためには、“コンテキスト”の明確な定義が必要となる。

コンテキストウェアという用語は、Schilitらの文献[53]で初めて使用された。文献[53]では、ユーザの近くの人物や物体の位置情報および識別情報をコンテキストとして扱っている。Ryanら[54]は、ユーザの位置情報および環境、識別情報、時刻をコンテキストとしている。Deyら[55]は、環境中の人物や物体はもちろん、ユーザの感情や注目の対象、位置、向いている方向、日時などもコンテキストとして扱っている。さらに、コンテキストを「実体(人物や場所、物体など)の状況の特徴付けるために用いられる情報」と定義している。

コンテキストに基づいた処理を実行するためには、何らかの方法によりコンテキストを表現する必要がある。Strangら[56]は、コンテキストの表現方式を、コンテキストモデルという概念に基づいて、以下のように分類している。

- Key-Value models
- Markup scheme models
- Graphical models
- Object oriented models
- Logic based models
- Ontology based models

これらは、あらかじめ人手(サービス設計者やユーザ)により、提供するサービスとコンテキスト情報とを関連付けておくことでモデル化を行う。このモデルを定義しておくことにより、コンテキストウェアシステムはユーザのコンテキストに応じたサービスを実行することが可能となる。例えば、「観光地において、

ユーザが観光施設 X の近くに来たら，観光施設 X の情報を提示する」といったルールをモデルとして定義しておく．この例のように，人手でモデルを作成するため，モデル作成者の意図どおりのサービス提供を行うことができる．

しかし，本論文で目的とするコンテキスト依存型情報推薦システムは，アイテムに対するユーザの嗜好に対応するものである．ユーザの嗜好は個人によって変わってくるため，事前に提供アイテムとコンテキスト情報とを関連付けておくことは容易でない．また，膨大に存在するアイテムに対し，ユーザ自身が逐一コンテキストと関連付けることも，ユーザにとって非常に負担になり，現実的でない．

そこで，本論文では機械学習手法により，ユーザコンテキストに依存したユーザ嗜好を動的に学習する方式を採用する．こうすることにより，ユーザが推薦システムを利用していくうちに，個人に特化したモデルが作成されるようになる．結果として，個々のユーザに対し，ユーザコンテキストに応じた適切な推薦を行うことが可能となる．

3. ユーザコンテキストの定義とコンテキスト依存型情報推薦システム

3.1 緒言

本章では、コンテキスト依存型情報推薦システムについて述べる。コンテキスト依存型情報推薦システムを設計する際には、コンテキストを明確に定義する必要がある。そこで、3.2節では、まず先行研究において扱われているコンテキストの定義を示し、本研究におけるユーザコンテキストの定義について述べる。3.3節では、本研究で提案するコンテキスト依存型情報推薦システムの全体像を提示し、次章以降で述べる各推薦方式の位置付けを示す。

3.2 ユーザコンテキストの定義

2.2節で述べた、コンテキストウェアシステムの研究分野においては、今現在コンテキストの確固たる定義がなく、多くの研究者によりそれぞれ独自の観点からコンテキストの定義が試みられている [57]。

Schilitら [53] は、コンテキストを次の三つのカテゴリに分類している。

Computing context ネットワークの接続状態や通信コスト、通信帯域幅、近くの資源（プリンタやディスプレイ、ワークステーションなど）など。

User context ユーザプロフィールやユーザの位置、近くの人々、その人との関係など。

Physical context 照明や騒音レベル、交通状況、気温など。

さらに、Chenら [57] は、時間情報もコンテキストとして重要であると考え、上記の三つのカテゴリに加え、下記のカテゴリを定義している。

Time context 日時や曜日、月、季節など。

Schmidtら [58][59] は、コンテキストを「ユーザおよびIT機器の状態に関する知識（含、環境や状況、位置）」と定義している。Deyら [60][55] は、コンテキス

表 1 ユーザコンテキストの例（繁華街における回遊情報を推薦対象コンテンツとした場合）

タイプ	例
推薦時コンテキスト	月, 曜日, 時刻, 天気, 気温, 湿度, 休日/非休日, 休日前/非休日前, 予算, 余裕時間 同伴者のタイプ, 同伴者の人数など
過去コンテキスト	～時に食事をした, ～時に買物をした, など
未来コンテキスト	～時に食事をする予定, ～時に買物をする予定, など

トを「実体の状況の特徴付けるために用いられる情報。実体とは、ユーザとアプリケーションとのインタラクションに関連すると考えられる、人や場所、物体のことをいう。さらに実体にはユーザ自身やアプリケーション自体をも含まれる。」と定義している。

以上のように、既往の研究では、ユーザ自身がおかれている環境や状況、周辺の各種デバイスの状態、時間情報などがコンテキストとして定義されている。

本論文では、情報選択過程におけるユーザの嗜好がさまざまな状況により影響を受ける点に着眼している。したがって、単に状況の特徴付けるための情報ということだけでなく、ユーザの嗜好に影響を及ぼすか否かといった観点から、コンテキストを定義する必要がある。そこで本論文では、「情報選択過程におけるユーザの嗜好に影響を及ぼす、ユーザのおかれている状況」をユーザコンテキストと定義する。特に、ユーザに関わってくる状況であるため、本論文ではユーザコンテキストという呼称を用いる。

具体的に、どの情報をユーザコンテキストとして扱うかは、推薦対象とするコンテンツに応じて大きく異なってくるため、一概に決定することは難しい。情報推薦システムを設計する際に、設計者が推薦対象コンテンツに応じてユーザコンテキストを決定する必要がある。表 1 に、繁華街における回遊情報を推薦対象コンテンツとしたときの、ユーザコンテキストの例を示す。

表 1 に示すように、ユーザコンテキストを時間軸の観点から整理すると、大きく次の二つのカテゴリに分類される。

推薦時コンテキスト ユーザがシステムから推薦情報を受ける時点の状況。

過去/未来コンテキスト ユーザがシステムから推薦情報を受ける時点を基準にし

たときの，過去/未来にユーザが取った/取る行動．

本論文では，第4章および第5章において，推薦時コンテキストを扱ったコンテキスト依存型情報推薦方式を，第6章において，過去/未来コンテキストを含めたコンテキスト依存型情報推薦方式を，それぞれ提案する．

3.3 コンテキスト依存型情報推薦システム

図1は，本論文で提案するコンテキスト依存型情報推薦システムの構成図である．ユーザは，携帯電話やPDAなどのモバイル機器を用いて，本システムを利用する．ここでは，推薦システムを利用するユーザを対象ユーザと表記する．

本システムによる推薦処理の流れは下記のとおりである（各手順の番号は図1中の番号に対応している）．

- (1) 対象ユーザの現在おかれているユーザコンテキストを取得する．
- (2) 対象ユーザのコンテキスト依存型ユーザ嗜好モデルに基づき，推薦対象のアイテムデータベースの中から推薦時コンテキストに合ったアイテム集合を選定し，推薦候補アイテム集合とする．
- (3) 推薦候補アイテム集合に対し，推薦時コンテキストに応じてランキングし，推薦リストを作成する．
- (4) 推薦候補アイテム集合に対し，過去/未来のコンテキストに応じてランキングし，推薦リストを作成する．
- (5) 推薦リストをユーザに提示する．

本論文において，第4章での提案方式については上記手順の(2)に，第5章での提案方式については上記手順の(3)に，第6章での提案方式については上記手順の(4)に，それぞれ相当する．

なお，(1)のユーザコンテキストの取得方法に関しては，これまでもさまざまな文献[58]において研究されているため，本論文では議論の対象外とする．本

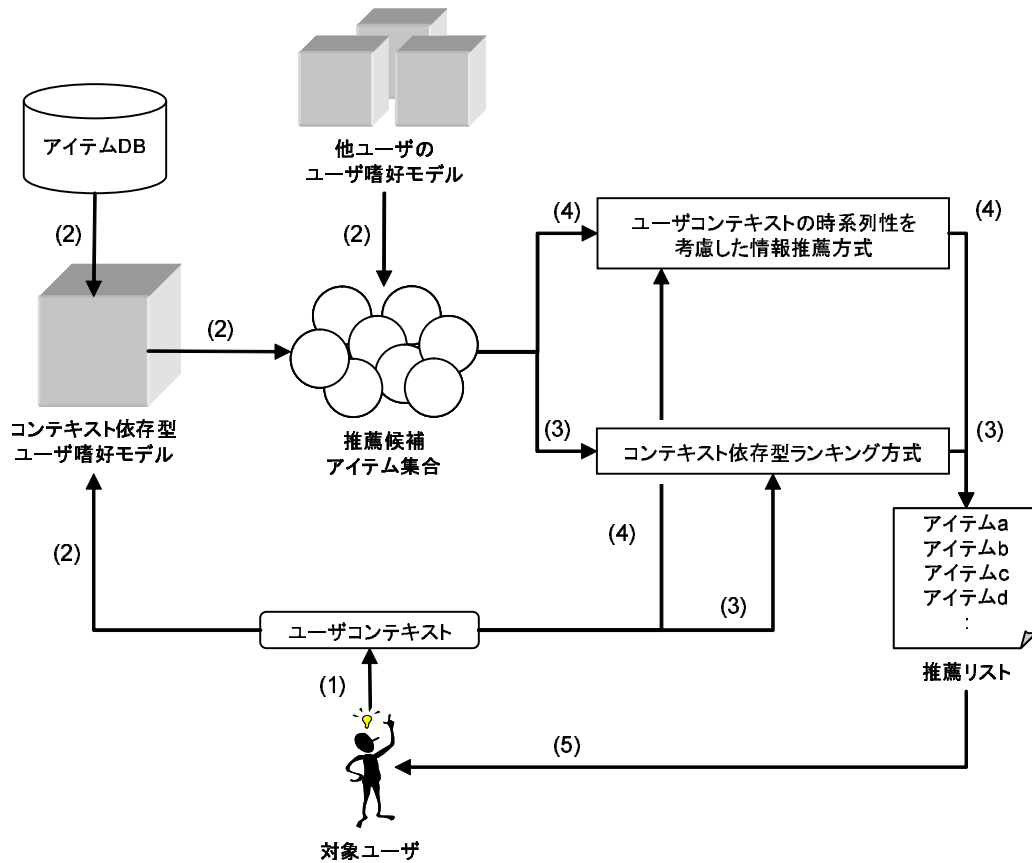


図 1 本論文で提案する情報推薦システムの構成図

論文では、ユーザコンテキストはモバイル機器に装備されているセンサやユーザ入力などにより取得できるものとして扱う。

4. コンテキスト依存型ユーザ嗜好モデルに基づく情報推薦方式

4.1 緒言

従来の情報推薦技術では，ユーザのアイテムに対する好みに関するデータを蓄積し，このデータに基づいてユーザ嗜好モデルを作成する．しかし，ユーザの嗜好はユーザコンテキストに応じて多様に変化するため，アイテムに対する好みに関するデータを入力とするだけでは適切なユーザ嗜好を表現することはできない．例えば，あるユーザにとって「夜に友人と行く焼肉店には満足」であったとしても「朝に一人でいる場合の焼肉店には不満足」であることも考えられる．また，「快晴の日にオープンカフェに行くことが好き」というユーザがいたとしても「雨の日に行くオープンカフェは嫌い」ということも考えられる．したがって，一概に焼肉店やオープンカフェといった，アイテムそのものに対する好みに関するデータからだけでは，ユーザコンテキストに応じて変化するユーザ嗜好に対応したモデルを作成することはできない．

そこで，本章では，ユーザコンテキストも含めたユーザ嗜好モデルである，コンテキスト依存型ユーザ嗜好モデルを提案する．また提案モデルを利用した，コンテキスト依存型情報推薦方式として，以下の二つの情報推薦方式を提案する．

- (a) コンテキスト依存型情報フィルタリング方式 (Context-Aware Information Filtering Method; C-IF)
- (b) コンテキスト依存型協調フィルタリング方式 (Context-Aware Collaborative Filtering Method; C-CF)

さらに，本章では，検証実験に基づき，提案のモデル化方式および情報推薦方式の評価も併せて行う．本章の成果の概要は以下のとおりである．

- 主なクラス分類手法との比較に基づき，ユーザ嗜好のモデル化方式として Support Vector Machine (SVM) [18] を用いることの妥当性の評価を行った．

- より効果的な情報推薦を実現するために，二つの情報推薦方式（上記の (a) および (b)）の有効性や利用場面に応じた特長の違いを分析した．
- アイテムおよびコンテキストの特徴パラメタの最適化手法について検討した．

以下，本章の構成を示す．4.2 節では，ユーザ嗜好のモデル化に用いるクラス分類手法として，SVM について説明する．4.3 節では，本章での提案方式である，コンテキスト依存型ユーザ嗜好モデル化方式について説明し，4.4 節では，このモデル化方式を適用したコンテキスト依存型情報推薦方式について説明する．4.5 節および 4.6 節，4.7 節では，それぞれ検証実験に基づき，提案方式の評価および考察を行う．最後に，4.8 節で本章をまとめる．

4.2 基礎的事項

本章では，ユーザの経験により蓄積された学習データ集合をユーザにとって満足なデータと不満足なデータに分類するという観点から，クラス分類手法を用いる．クラス分類手法により未知データに対してユーザに推薦すべきか否かの判定を行う．

クラス分類手法には，Support Vector Machine (SVM)[18] やニューラルネットワーク [24]， k 最近傍法 (k NN) [61]，決定木 [22][23]，ベイズ分類器 [62] などがある．

このうち本章では以下のような利点を持つ SVM を用いる．

- 少ない学習データから識別関数を構成できる．
- 膨大な次元数を持つ学習データにも対応可能．
- 解が一意に定まる．
- 最適化すべき学習パラメタが少ない．

SVM は 2 クラスの分類手法の一つであり，1995 年に Vapnik ら [18] によって提案された．SVM は主に画像認識や自然言語処理などの分野で利用されている．

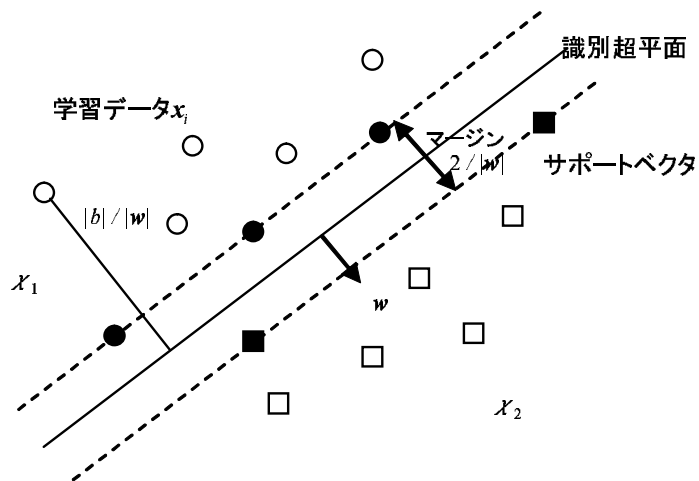


図 2 SVM の特徴空間

図 2 のように学習データが分布する空間を SVM の特徴空間という．図 2 中の“○”，“●”，“△”，“▲”は学習データを表現し，それぞれの形状は正データ，負データを表す．SVM では，学習データ集合に対する複数の識別境界候補の中で，図 2 に示すようなマージンを最大にする識別超平面を求める．このマージン最大化が SVM の特徴であり，これにより解が一意に求められる．図 2 の“●”および“▲”はサポートベクタとよばれ，実際にはこのサポートベクタだけから識別超平面が構成される．

4.3 コンテキスト依存型ユーザ嗜好モデル化方式

本節では，コンテキスト依存型ユーザ嗜好モデル化方式について説明する．本章では，4.2 節で述べた，SVM[18] を用いる．情報推薦においては，未知のデータに対して，ユーザの嗜好をいかに予測できるかが重要であるため，本章では，この高い汎化性能を持つ SVM をユーザ嗜好のモデル化方式に適用した．

4.3.1 項では，単純なモデル化方式による問題点を指摘した上で，それを解決する提案のモデル化方式のアイデアを説明する．4.3.2 項では，4.4 節で提案するコンテキスト依存型協調フィルタリングを実装するための事前準備として，ユーザコンテキストを考慮したユーザ類似度の算出方法について説明する．

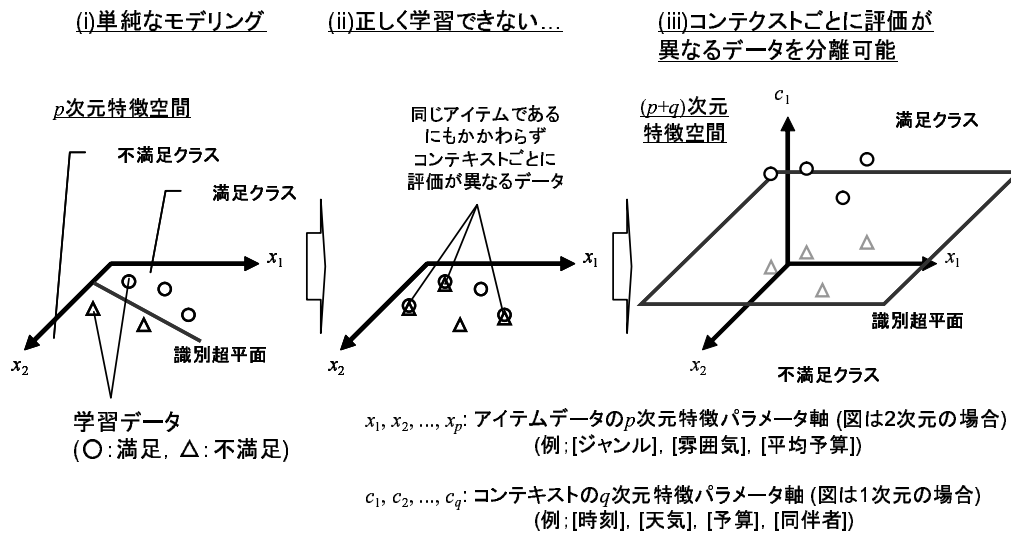


図 3 コンテキスト依存型ユーザ嗜好モデル化方式

4.3.1 ユーザ嗜好のモデル化手順

まず, SVM を用いた場合の, 単純なユーザ嗜好のモデル化の手順を以下に示す (図 3(i) 参照) .

- (1) アイテムデータを p 次元の特徴パラメタ ([ジャンル] や [雰囲気], [平均予算] など) の形式で表現する .
- (2) ユーザはアイテムデータ一つに対して好みの評価 (“満足” ・ “不満足”) を与える .
- (3) 評価付けされたアイテムデータ集合を学習データ集合として p 次元特徴空間上に配置する .
- (4) 学習データ集合を満足クラスと不満足クラスとに分類する識別超平面を構成する .

上記手順により, アイテムデータに対するユーザの嗜好をモデル化することができるように見える . しかし, 人の嗜好は, その人のおかれているコンテキストにより変化するものである . 例えば, 「同じ焼肉店であっても, 夜に友人と行くと

きには満足だが、朝に一人で行くときには不満足」といったように、同じアイテムであるにもかかわらず、コンテキストごとに評価が異なるということが実際にはあり得る (図 3(ii) 参照)。このように、その時々コンテキストに応じて多様に変化するようなユーザの嗜好をモデル化することは容易ではなく、上記手順ではこれを適切にモデル化することはできない。

この問題に対して、本章ではデータの表現方法に着目する。従来方法では、データの特徴量を表現するものとしてアイテムの特徴を表すパラメタだけしか扱わなかった。これに対し、本章では、このデータの特徴量表現に対してコンテキストの特徴を表すパラメタも含めて統合する。このようにデータの特徴量としてコンテキストの特徴パラメタも含めることにより、例えば「夜に友人と行くときの焼肉店」と「朝に一人で行くときの焼肉店」とを異なるデータとして扱うことができる。これはつまり、「同じアイテムデータであっても、評価時のコンテキストが変われば異なるデータである。」という考えに基づくものである。これに合わせて、用いる SVM の特徴空間も、コンテキストの特徴パラメタの次元数だけ拡張する。

その具体的なモデル化の手順を以下に示す (図 3(iii) 参照)。

- (1) アイテムデータを p 次元の特徴パラメタ、コンテキストを q 次元の特徴パラメタ ([時刻] や [天気], [予算], [同伴者] など) の形式で、それぞれ表現する。
- (2) ユーザはアイテムデータに対して一つ一つ、その時のコンテキストにおける好みの評価 (“満足”・“不満足”) を与える。
- (3) 評価付けされたアイテムデータの特徴パラメタとその評価時のコンテキストの特徴パラメタとを統合したデータ集合を、学習データ集合として $(p+q)$ 次元特徴空間上に配置する。
- (4) 学習データ集合を満足クラスと不満足クラスとに分類する識別超平面を構成する。

本章ではこのモデル化方式を、コンテキスト依存型ユーザ嗜好モデル化方式とよぶ。CP モデリングにより、コンテキストに応じて変化するユーザ嗜好が適切に

学習されたユーザ嗜好モデル，コンテキスト依存型ユーザ嗜好モデル (Context-Aware User Preference Model; 以降，CP モデルと称す) を構築することができる．CP モデルの特長は，さまざまなアイテムの特徴およびコンテキストの特徴に対する複雑なユーザ嗜好を唯一つのシンプルな特徴空間で表現しているところにある．

4.3.2 ユーザ嗜好モデルに基づくユーザ類似度の算出手順

従来の協調フィルタリングでは，ユーザ同士の嗜好の類似度を算出するための主な方法として，同じアイテムに対して与えた評価値の相関係数に基づく方法がある．例えばユーザ A がカフェ X を好むとし，ある別のユーザ B もそのカフェ X を好むとしたとき，ユーザ A とユーザ B が嗜好が類似しているユーザ (類似ユーザ) として扱われる．しかし，前節で述べたことと同様，同じアイテムに対しても，その時のコンテキストにより評価が異なる．例えば，ユーザ A が「朝に一人で行くときには，カフェ X は満足」としているのに対し，ユーザ B が「夕方に恋人と行くときには，カフェ X は満足」としていた場合，ユーザ B は「朝に一人で行くときには，カフェ X は不満足」と評価する可能性もある．この場合，ユーザ A とユーザ B は嗜好が類似しているとはいえない．つまり，ユーザの評価時のコンテキストを考慮しないと，ユーザの嗜好の類似度を適切に算出することはできない．

この問題に対し，我々は，互いのユーザの CP モデルの類似度に基づいてユーザ類似度を算出する，という方法を取る．前節で説明したように，CP モデルには，そもそもユーザの評価時のコンテキストの情報も含まれているため，このモデルの類似度に基づいてユーザ類似度を算出することにより，評価時のコンテキストも含めて適切にユーザの嗜好の類似度を算出することができる．

その具体的なユーザ類似度の算出手順を以下に示す．図 4 のように，あらかじめ CP モデルが構築された 2 人のユーザ，ユーザ A とユーザ B を対象に，ユーザ類似度の算出方法を説明する．

- (1) 各ユーザの学習データ集合を互いの CP モデル上に交換する．

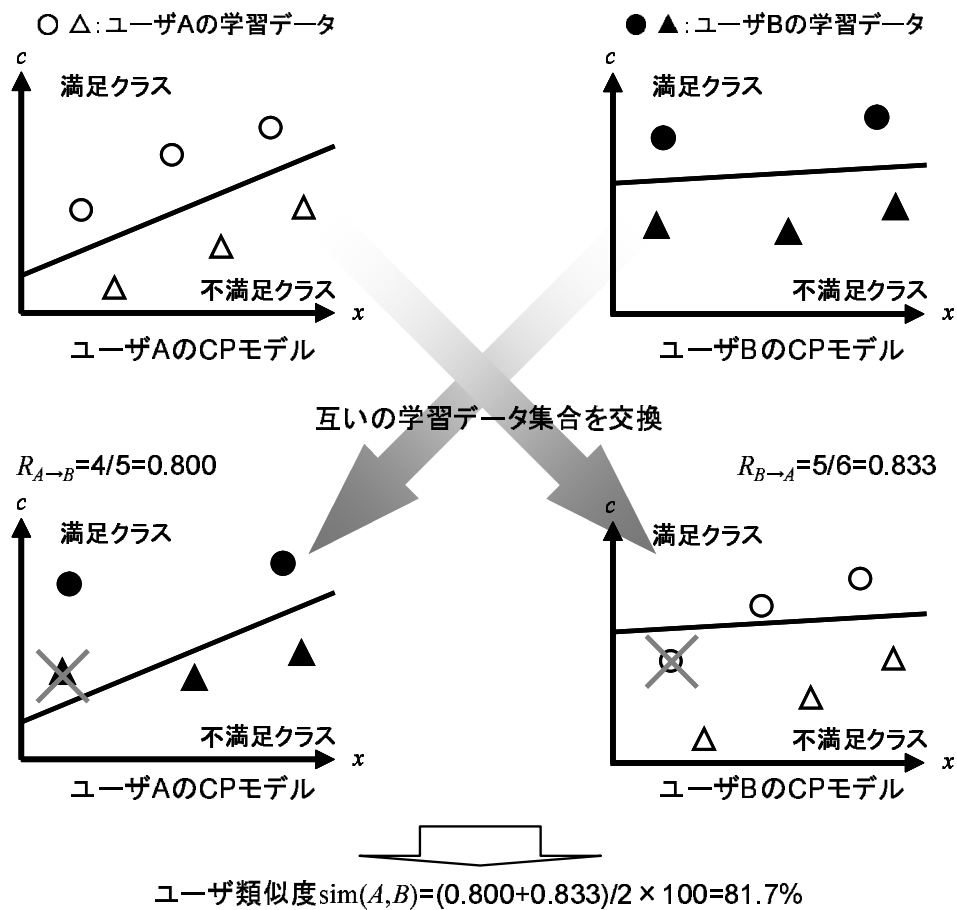


図4 CPモデルに基づくユーザ類似度の算出

- (2) ユーザAのCPモデルにより、ユーザBの学習データ一つ一つに対して属するクラス(“満足”・“不満足”)を判定する。
- (3) そのユーザAのモデルにより判定されたクラスとユーザBの評価値とが一致したデータ数($M_{A \rightarrow B}$)を数える(図4の例では $M_{A \rightarrow B} = 4$)。
- (4) ユーザBの全学習データ数(N_B)に対する一致データ数の割合($R_{A \rightarrow B}$)を求める(図4の例では $R_{A \rightarrow B} = 4/5 = 0.800$)。
- (5) ユーザAの学習データ集合に対しても同様に行い、 $R_{B \rightarrow A}$ を求める(図4の例では $R_{B \rightarrow A} = 5/6 = 0.833$)。

(6) $R_{A \rightarrow B}$ と $R_{B \rightarrow A}$ との平均値をユーザ類似度 ($\text{sim}(A, B)$) とする (図 4 の例では $\text{sim}(A, B) = 81.7\%$) .

以上の算出方法を一般化すると, ユーザ u とユーザ v とのユーザ類似度 $\text{sim}(u, v)$ は, 式 (1) により求められる .

$$\text{sim}(u, v) = \frac{1}{2} \left(\frac{M_{v \rightarrow u}}{N_u} + \frac{M_{u \rightarrow v}}{N_v} \right) \times 100(\%) \quad (1)$$

このユーザ類似度が高いユーザ同士を類似ユーザとして扱う .

互いのユーザのアイテムに対する評価値の相関関係に基づく従来の算出方法では, 両ユーザがいくつかの同じアイテムに対して評価を与えていることが前提となる . これに対して, CP モデルに基づくこの方法では, 必ずしもユーザ間で同じアイテムに対して評価付けされているという必要はない .

4.4 コンテキスト依存型ユーザ嗜好モデルに基づく情報推薦方式

本節では, 前節で説明した CP モデルに基づいたコンテキスト依存型情報推薦方式として, 以下の二つの情報推薦方式について説明する .

- (a) コンテキスト依存型情報フィルタリング方式 (Context-Aware Information Filtering Method; C-IF)
- (b) コンテキスト依存型協調フィルタリング方式 (Context-Aware Collaborative Filtering Method; C-CF)

本節では, 飲食店情報推薦を例として, これらの情報推薦方式について説明する .

4.4.1 コンテキスト依存型情報フィルタリング方式 (C-IF)

コンテキスト依存型情報フィルタリング方式 (以降, C-IF と称す) は, 対象ユーザの CP モデルに基づいて, アイテムデータベースから, ユーザのおかれているコンテキストにおけるユーザの嗜好に合致するアイテムを検索し, そのアイテムをユーザに推薦しようというものである .

以下, C-IF による推薦手順を示す (図 5 参照) .

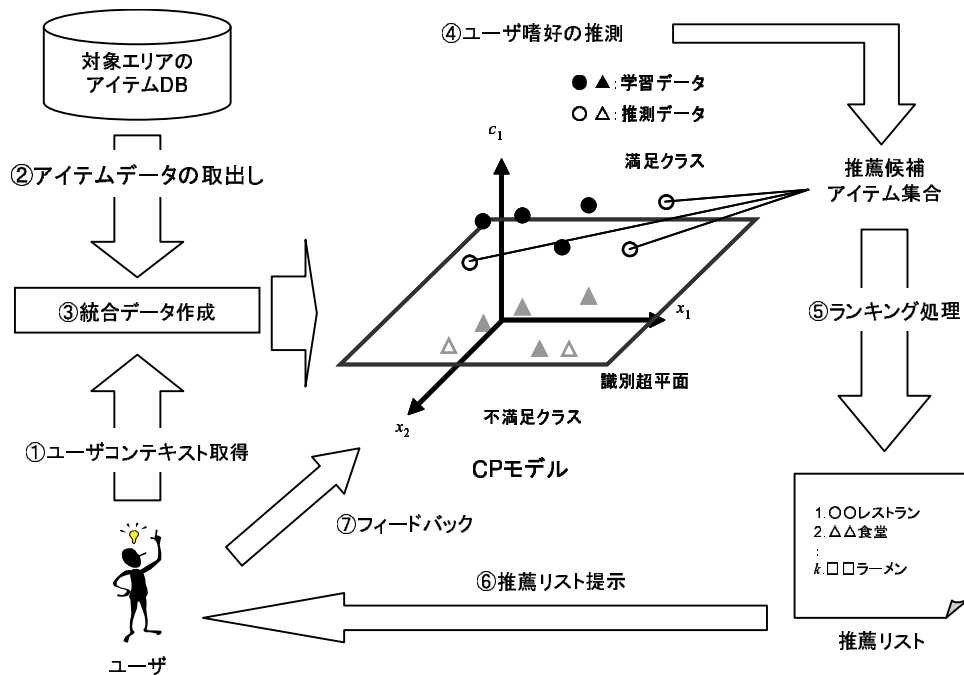


図 5 コンテキスト依存型情報フィルタリング方式 (C-IF) による推薦手順

- (1) ユーザコンテキストを自動的に、もしくはユーザによる明示的な入力により取得する。
- (2) 対象エリア (ユーザが現在位置しているエリア) のアイテムデータベースから、ユーザの現在位置に最も近い n 件のアイテムデータを取り出す。
- (3) 取り出したアイテムデータ一つ一つの特徴パラメタとユーザコンテキストの特徴パラメタとを統合したデータ集合を作成する。
- (4) 統合したデータ集合を CP モデル上に配置し、この CP モデルに基づいて、各データに対するユーザの嗜好 (“満足”・“不満足”) を推測する。
- (5) ユーザの現在おかれているコンテキストにおいて満足されると推測されたデータ集合を推薦候補のアイテムデータ集合とし、このデータ集合に対してランク付けを行う。

- (6) そのランキングにしたがって、上位 k 件のアイテムデータを推薦リストとして、ユーザに提示する。
- (7) 推薦データに対するユーザのフィードバックに基づいて新たな学習データを追加する。

4.4.2 コンテキスト依存型協調フィルタリング方式 (C-CF)

コンテキスト依存型協調フィルタリング方式 (以降、C-CF と称す) では、ユーザ同士の CP モデルの類似度に基づいて、類似ユーザを探し出し、対象ユーザのコンテキスト近傍において、その類似ユーザが満足と評価しているアイテムを対象ユーザに推薦しようというものである。

以下、C-CF による推薦手順を示す (図 6 参照)。ここでは、まず 4.3.2 項で説明したユーザ類似度の算出手順にしたがって、あらかじめ対象エリアに存在する類似ユーザを探索しておき、ユーザ類似度が最も高い m 人のユーザの CP モデルを参照することにする。図 6 では、対象ユーザをユーザ A 、そのユーザ A の類似ユーザとして、ユーザ B およびユーザ C が発見されているとする。

- (1) 対象ユーザのコンテキストを自動的に、もしくはユーザによる明示的な入力により取得する。
- (2) m 人の類似ユーザの CP モデルに基づいて、対象ユーザのおかれているコンテキストの近傍において、満足と評価されているアイテムデータ集合を対象ユーザの CP モデル上に配置する。
- (3) 対象ユーザの CP モデルに基づいて、配置された各アイテムデータに対するユーザの嗜好 (“満足”・“不満足”) を推測する。
- (4) ユーザの現在おかれているコンテキストにおいて満足されると推測されたデータ集合を推薦候補のアイテムデータ集合とし、このデータ集合に対してランク付けを行う。
- (5) そのランキングにしたがって、上位 k 件のアイテムデータを推薦リストとして、ユーザに提示する。

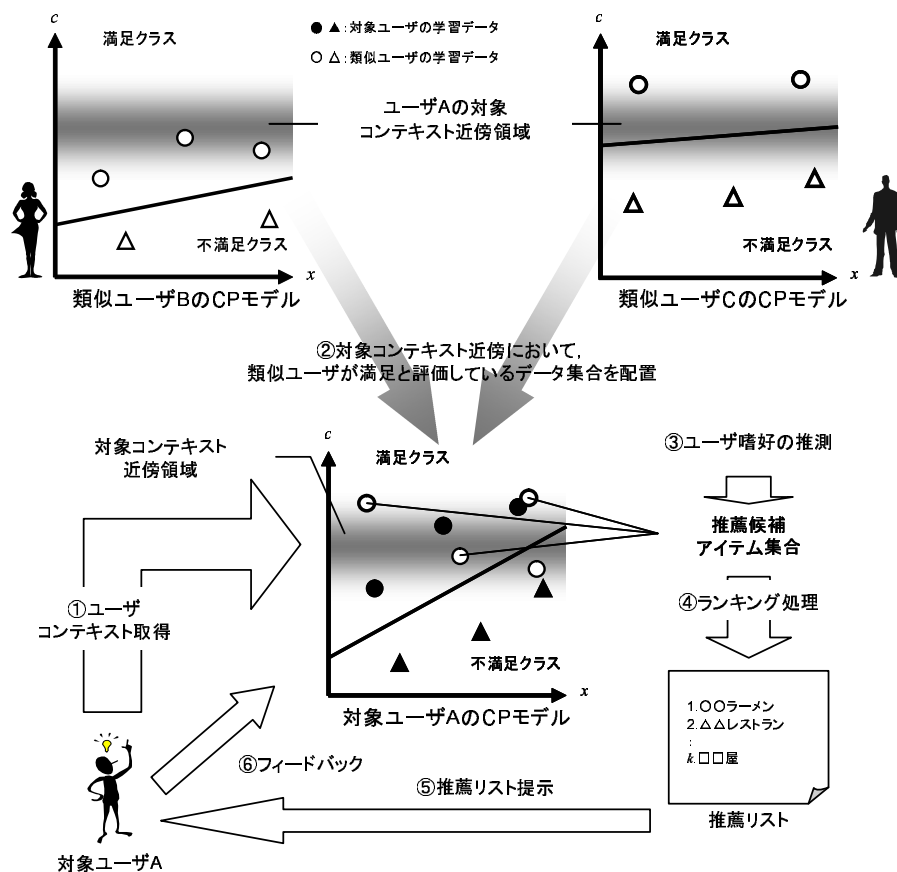


図 6 コンテキスト依存型協調フィルタリング方式 (C-CF) による推薦手順

- (6) 推薦データに対するユーザのフィードバックに基づいて新たな学習データを追加する。

4.5 ユーザ嗜好モデル化方式としてSVMを適用することの妥当性の検証

情報推薦システムは、ユーザが把握しきれていないアイテムを適切に推薦することに価値がある。したがって、情報推薦方式に対して適用するモデル化方式は、既知 (学習済み) データに対して正しく識別できる能力も当然必要であるが、む

しる未知 (未学習) データに対して高い汎化性能を示すことが望ましい。本章では、一般に汎化性能に優れているといわれている理由から、ユーザ嗜好をモデル化するためのクラス分類手法としてSVM[18] を用いている。

SVM の他には、主なクラス分類手法として、Neural Network (NN) [24], k -Nearest Neighbor (k NN) [61], Decision Tree (DecTree) [22][23], Bayesian Filtering (BF) [62] が挙げられる。本実験では、汎化性能の観点から、これら五つの手法を比較対象とした定量評価実験を行った。本節では、この定量評価実験の結果を示し、それに対する考察を行う。

4.5.1 実験方法

データセット

まず飲食店情報推薦を対象とした実データセットを用いて、各クラス分類手法の汎化性能について比較評価を行う。また、データセットの特性の違いに対する汎用性を分析するために、次元数およびデータ数が異なる人工データセットに対しても比較評価を行う。

まず、飲食店情報推薦を対象とした、被験者の実データセットを用いる。実データセットは、飲食店データの特徴パラメタ 17 次元とコンテキストの特徴パラメタ 14 次元の合計 31 次元から構成される。データ数は 100 個であり、一つ一つのデータに対して被験者による評価値 (“満足”・“不満足”) が与えられている。

人工データセットとして、表 2 に示す五つのデータセットを用いる。各データセットは、LIBSVM のサイト [63][64] で提供されている 2 クラス用のデータセットのうち、それぞれ表 2 に示すようにデータ数および次元数が異なるように選択した。

評価指標

性能評価のための指標として識別性能を用いる。まず、学習データ集合に基づいて各クラス分類手法による識別モデルを構築する。つづいて、この識別モデルによりテストデータ集合に対して識別判定を行う。このとき、そのテストデータ集合に対してどの程度正しく識別できたかを示す指標が識別性能であり、式 (2)

表 2 クラス分類手法の性能分析に用いる人工データセット

データセット名	データ数	次元数
fourclass	862	2
heart	270	13
australian	690	14
german	1000	24
sonar	208	60

で表される．

$$\text{識別性能} = \frac{\text{正しく識別されたデータ数}}{\text{総テストデータ数}} \times 100(\%) \quad (2)$$

本実験では，未知データに対する識別性能すなわち汎化性能について検証する．これは，データセットの一部のデータを学習データとし，残りのデータをテストデータとしたときの識別性能である．具体的には，5-fold cross-validation により，総データ数の 4/5 を学習データ，残りの 1/5 をテストデータとし，この 5 通りの組合せに対する平均を取っている．これは，学習データおよびテストデータの選び方により生じるデータ分布の偏りをなくするためである．

なお，各クラス分類手法による識別性能の取得には，データマイニングツール WEKA[65] を用いた．

4.5.2 実験結果と考察

汎化性能に関する考察

図 7 および図 8 に実データセットおよび人工データセットを対象としたときの未知データに対する識別性能を示す．

この結果から，実データセットに対して，NN, k NN, DecTree, BF においては，未知データに対する識別性能が 60% 程度であったのに対し，SVM は 70% 近い性能を示した．また，五つの人工データセットいずれに対しても，未知データに対して SVM が比較的高い識別性能を示すことを確認した．

実データセットを対象としたときには，SVM は他の各手法に対して比較的高い汎化性能を示した．したがって，飲食店を対象コンテンツとした情報推薦システムにおいては，SVM を情報推薦方式に対して適用することは有用性があると

いえる．

今回は，実データセットとして飲食店情報推薦を対象としたものだけを扱ったが，データ数および次元数が異なるさまざまな人工データセットに対しても，比較的高い汎化性能を示したSVMは，データセットの特性にとらわれず汎用性があるといえる．したがって，推薦対象コンテンツを音楽や映画などに変えた場合でも，汎用的に有用性を示すことが期待できる．

特徴パラメタの次元数と学習データ数の関係に関する考察

一般に，多次元特徴ベクトルを用いた学習問題では，入力次元数の増大とともに，必要とされる学習データ数が指数関数的に増えていく（次元の呪い [61]）．したがって，モデルの識別性能を高めるためには，次元数に対して十分な数の学習データを用意するか，特徴選択手法などを用いて次元数を減少させなくてはならない．

これに対し，SVMは，入力次元数が増加しても，モデルの複雑さは増加しない，といった特徴を持つ [66]．平ら [67] は，SVMをテキスト分類へ応用したとき，入力次元数と学習精度の関係を調べている．この結果，SVMでは入力次元数が増加しても，精度を落とすことなく，学習できることを示している．

以上のように，高次元の特徴パラメタを持つ学習データにも対応可能であり，かつ少ない学習データからでも高い汎化性能を示すことができるという特徴を持つことから，SVMを，ユーザ嗜好モデル化方式として適用することは妥当であるといえる．

4.6 被験者満足度による推薦方式の評価

本節では，4.4節で説明した推薦方式の評価を行う．具体的には，飲食店を推薦対象コンテンツとしたとき，提案方式による推薦に対する被験者の満足度を用いて推薦方式の有用性を評価する．

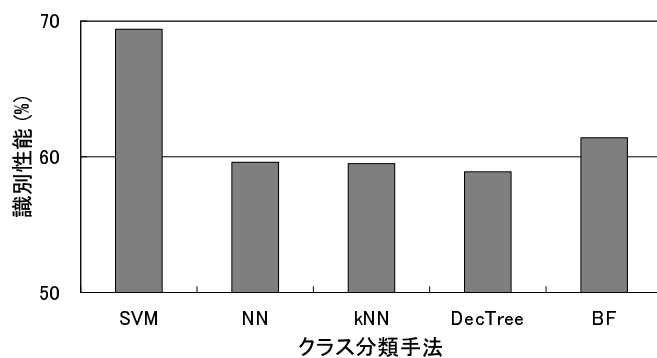


図 7 実データセットを対象としたときの未知データに対する識別性能の比較

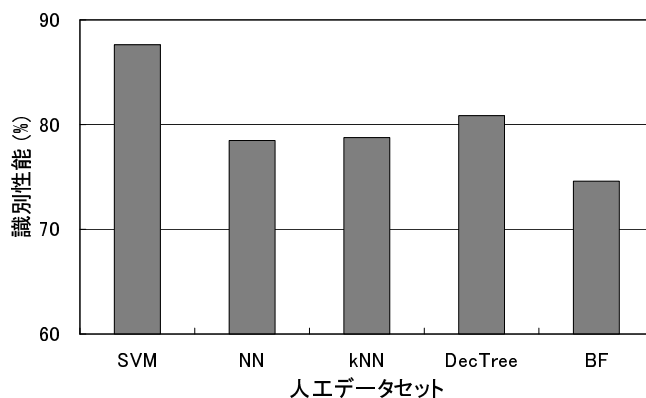


図 8 人工データセットを対象としたときの未知データに対する識別性能の比較

4.6.1 実験条件

データセット

本実験では、飲食店情報を対象としたデータセットを用いた。ここでは、一般向けグルメサイトである『Yahoo!グルメ [68]』を参考に、表 3 に示すエリアを対象に選択した『Yahoo!グルメ』に登録されているデータ件数のうち、本実験で用いることができる有効なデータだけを用いる。その有効データ数は表 3 に示すとおりである。さらに、『Yahoo!グルメ』を参考に、表 4 に示す 28 次元の飲食店特徴パラメタを設定した。

表 3 本実験で用いた飲食店データセット

エリア	登録件数	有効データ数
I 心齋橋、堀江	2069	650
II 天王寺、阿倍野	416	121
III 梅田、中津	1508	445

また、コンテキスト特徴パラメタについても、同様に表 5 に示す 24 次元の特徴パラメタを設定した。これらは、ユーザの飲食店情報の選択に対する意思決定に影響を及ぼすという観点から、著者が主観的に決めたものである。なお、これら各特徴パラメタの選定においては、特徴ベクトル次元の直交性確保の観点から、できる限り独立関係になるよう心がけた。

被験者と評価指標

表 5 に示したコンテキスト特徴パラメタの組合せにより表現されたユーザコンテキストにおいて、飲食店データに対する評価（“満足”・“不満足”）を被験者に行ってもらった。被験者は、22 歳から 35 歳の 9 名（被験者 a, b, ..., i）である。いずれの被験者も、表 3 に示したエリアには何度か訪れたことがあり、これらのエリアの雰囲気などについてはある程度把握している。このように被験者が把握しているエリアを実験対象として選択したのは、研究室内で仮想的に実験を行うため、本実験を行うに当たって、できるだけ被験者にユーザコンテキストを想像しやすくするためである。

なお、本実験では、被験者満足度を用いてシステムの評価を行う。満足度とは、被験者が情報推薦に対してどの程度満足したかを表す指標であり、式 (3) により表される。

$$\text{満足度} = \frac{\text{被験者が満足した回数}}{\text{推薦回数}} \times 100(\%) \quad (3)$$

情報推薦分野の研究では、情報推薦技術の基礎評価を行うために、10 名程度の被験者による主観評価を行っているものが多い [27][69][70]。本章においても、提案方式の基礎評価を行うものであるため、上記のような実験条件とした。

表 4 本実験で用いた飲食店特徴パラメタ (28 次元)

パラメータ項目	パラメータ値
ジャンル (8)	[和食] / [洋食] / [中華] / [居酒屋] / [アジア] / [ラーメン] / [カフェ、スイーツ] / [その他]
予算 (昼夜) (2)	0~10000円
店の特徴 (11)	[個室] / [駐車場] / [禁煙席] / [カラオケ] / [深夜営業] / [食べ放題] / [ランチ] / [カード可] / [ライブ] / [テイクアウト] / [貸し切り可]
雰囲気 (7)	[デート向き] / [接待向き] / [宴会・パーティー向き] / [合コン向き] / [ファミリー向き] / [1人でも可] / [少人数向き]

表 5 本実験で用いたコンテキスト特徴パラメタ (24 次元)

パラメータ項目	パラメータ値
時間情報	月 (2) 1~12月 曜日 (2) 日~土 時刻 (2) 00:00~23:59
気象情報	天気 (4) [晴れ] / [曇り] / [雨] / [雪] 気温 (1) 0~40℃
ユーザ情報	休日 (2) [休日] / [休日前] 予算 (1) 0~10000円 余裕時間 (1) 0~4時間
同伴者情報	グループ (7) [一人] / [家族] / [恋人] / [友人・先輩・後輩] / [上司・部下] / [ビジネス] / [該当なし] 人数 (男女) (2) 0~10人

4.6.2 ランダム提示およびコンテキスト非依存の推薦方式との比較

本項では、C-IF および C-CF による情報推薦の有効性を、ランダム提示およびコンテキストに依存しない推薦方式との比較を行うことにより評価する。

実験方法

本実験で用いるデータセットの対象エリアは、表 3 に示すうち「心齋橋，堀江」エリアとした。

まず、あらかじめ下記のとおり各被験者のユーザ嗜好モデルを構築しておく。

- (1) 被験者には、各自「心齋橋，堀江」エリアにおいて、現実起こりうるコ

ンテキストを想像してもらい、あらかじめ10パターンのコンテキストを設定してもらう。

- (2) 被験者には、一つのコンテキストあたり、データセットの中からランダムに選択された10件の飲食店データに対して、“満足”・“不満足”の評価を与えてもらう。
- (3) この評価に基づいて、被験者一人あたり合計100個 (=10コンテキスト×10飲食店データ) の学習データを作成する。
- (4) 作成された学習データに基づいてコンテキスト依存/非依存型のユーザ嗜好モデルを構築する (4.3.1項参照)。
- (5) コンテキスト依存型ユーザ嗜好モデルに基づいて各被験者間のユーザ類似度を算出する (4.3.2項参照)。ここで、各被験者に対してユーザ類似度が最も高い3人の被験者を C-CF における類似ユーザとして扱う。

つづいて、以下の各方式で被験者に対して飲食店を推薦する。

- (i) ランダム提示：被験者の嗜好およびコンテキストに関係なく、ランダムに飲食店を提示する。
- (ii) コンテキスト非依存型推薦：被験者のコンテキストは考慮しないが、ある程度の嗜好に基づいて飲食店を推薦する。
- (iii) C-IF による推薦：コンテキストごとに変化する被験者の嗜好を考慮して飲食店を推薦する。
- (iv) C-CF による推薦：コンテキストも考慮した類似ユーザの嗜好に基づいて飲食店を推薦する。

被験者は、各方式により推薦された飲食店に対して、“満足”・“不満足”の評価を与える。この評価に基づいて、被験者満足度を算出する。

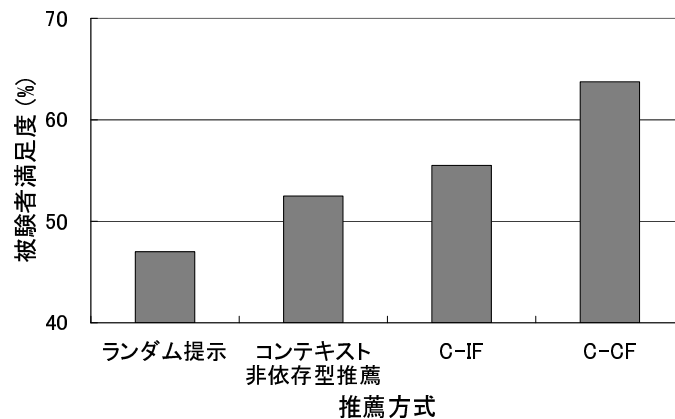


図 9 各推薦方式に対する被験者満足度の比較

実験結果と考察

図 9 に各方式に対する被験者満足度を示す．これは，9 名の被験者の満足度の平均を示している．C-CF に対する満足度が最も高く，次いで，C-IF に対する満足度が高かった．

ランダム提示は，そもそも提示する飲食店に根拠がないため，最も満足度が低かったということは自明の結果である．コンテキスト非依存型推薦は，コンテキストごとに変化するユーザの嗜好を正しくモデル化できていないことから，ときにコンテキストにふさわしくない飲食店が推薦されたことなどが満足度が低くなった原因であると考えられる．これに対して，C-IF は，コンテキストごとに変化するユーザの嗜好を正しくモデル化しているため，コンテキストに合った飲食店が適切に推薦されたため，比較的高い満足度が得られたといえる．さらに，C-CF は，コンテキストに適合し，かつ一度他のユーザが満足と評価した飲食店を推薦しているため，推薦情報に対する信頼度が高く，結果，最も高い満足度が得られたといえる．

以上のように，C-IF および C-CF の有効性を示した．なお，本実験では結果的に C-CF に対する満足度が最も高くなったが，これは，推薦対象エリアが固定であり，また各被験者がそのエリアにおいてあらかじめ学習データを作成しているという条件下での結果である．そもそも周りに類似ユーザが存在しない場合など，

システムの利用場面の違いにより，C-IF および C-CF の有効性も変わってくる可能性もある．次項では，これを踏まえて，推薦対象エリア別の C-IF および C-CF の特長の違いを検証する．

4.6.3 C-IF および C-CF の推薦対象エリア別の有効性の比較

本節では，C-IF および C-CF による有効性を推薦対象エリア別に比較することによって，各々の推薦方式の利用場面に応じた特長の違いを分析する．

実験方法

実験方法は以下のとおりである．

- (1) 9名の被験者を3名ずつの三つの被験者グループ ($A = \{a, b, c\}$, $B = \{d, e, f\}$, $C = \{g, h, i\}$) に分ける．
- (2) 各被験者グループはそれぞれ対応するエリア (グループ $A \rightarrow$ エリア I, グループ $B \rightarrow$ エリア II, グループ $C \rightarrow$ エリア III) で学習データを作成する．
- (3) 三つのエリアにおいて，C-IF および C-CF により各被験者に対して飲食店を推薦する．
- (4) 被験者は，推薦された飲食店に対する評価 (“満足”・“不満足”) を与える．

被験者一人あたりの学習データ数は100個 (=10 コンテキスト \times 10 飲食店データ) とする．推薦回数は，各被験者に対して，1 エリアあたり C-IF および C-CF 各 25 回 (=5 コンテキスト \times 5 飲食店データ) ずつの 50 回，3 エリアで合計 150 回とする．ここで，C-CF において用いる類似ユーザは，推薦対象エリアを学習エリアとしたグループに属する被験者すべてとする．例えば，被験者 a がエリア II を推薦対象エリアとしていた場合，そのエリアにおいて学習データを作成した被験者 d, e, f が，被験者 a の類似ユーザとなる．

なお，本実験では，被験者が学習データを作成したエリアを，その被験者にとって「既知エリア」とし，既知エリア以外のエリアを「未知エリア」とよぶ．例え

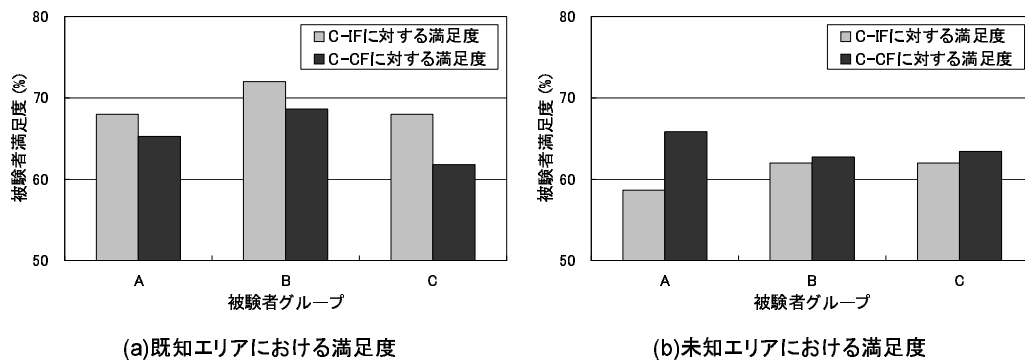


図 10 既知/未知エリアにおける C-IF および C-CF に対する満足度の比較

ば、被験者グループ A の場合、エリア I が既知エリアであり、エリア II および III が未知エリアである。

実験結果と考察

図 10 に、対象エリア別の被験者満足度を示す。(a) は既知エリアを推薦対象としたときの、(b) は未知エリアを推薦対象としたときの被験者満足度をそれぞれ示している。本実験より以下の結果が得られた。

- 既知エリアを推薦対象とした場合には、C-IF に対する満足度の方が高い。
- 未知エリアを推薦対象とした場合には、C-CF に対する満足度の方が高い。

この結果は、C-IF は対象ユーザ自身の嗜好が学習されたモデルに基づく推薦方式であるため、その学習データが十分に蓄積されたエリアを推薦対象としたときには、C-IF による情報推薦の方が有効であるということを示唆している。一方で、推薦対象エリアにおける学習データが乏しいときには、代わりにそのエリアにおいて十分な学習データをもつ他の類似ユーザのモデルに基づいた推薦方式、つまり C-CF による推薦方式の方が有効であるということも示唆している。

C-IF は対象ユーザの嗜好に忠実にしたがった、正確性を重視した推薦方式であるといえ、対して、C-CF は類似ユーザの嗜好や経験に基づくため、意外性を重視した推薦方式であるといえる。このような C-IF の正確性、C-CF の意外性といった各々の特長点を踏まえて、推薦対象エリアなどの利用場面に応じて、推薦

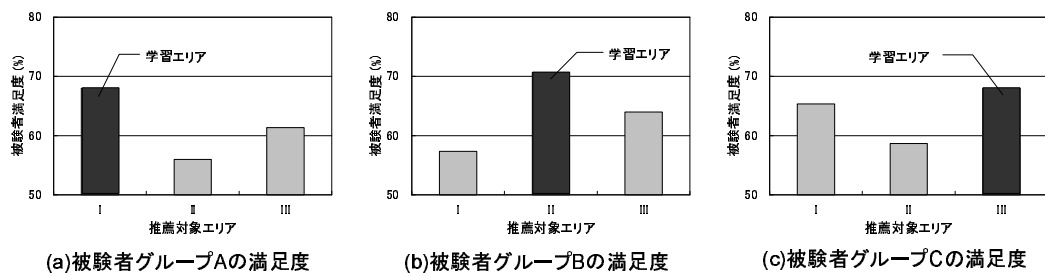


図 11 各被験者グループの推薦対象エリア別の C-IF に対する満足度の比較

方式に対する優先度を変えることにより，正確性と意外性を考慮した情報推薦が可能であるといえる。

4.6.4 推薦対象エリア特性に基づく C-IF の有効性の分析

さらに，C-IF による情報推薦に焦点をあてて，前項で得られた被験者満足度を被験者グループ別および推薦対象エリア別の傾向を分析する。

図 11 に C-IF に対する推薦対象エリア別の被験者満足度を示す。被験者グループ別にグラフを分けて示しており，(a)，(b)，(c) がそれぞれ被験者グループ A, B, C の被験者満足度に対応する。

この結果から，前節の結果で示したとおり，いずれの被験者グループも学習対象エリアにおいて高い満足度を示していることが分かる。被験者グループ A とグループ C について，より深くみると，被験者グループ A はエリア III を推薦対象としたときにも高い満足度が得られ，逆に被験者グループ C はエリア I を推薦対象としたときにも高い満足度が得られたことが分かる。エリア I 「心斎橋，堀江」およびエリア III 「梅田，中津」は，共に大阪都心部の歓楽街であり，店舗分布や利用者の活動内容などの地域特性が類似している [71]。この結果は，推薦対象エリアと学習エリアとの地域特性が類似している場合には，たとえ，ユーザにとって未知のエリアであったとしても，C-IF による情報推薦も有効に働くということを示唆している。

4.7 特徴パラメタの最適化に関する検討

既に説明したように、本実験では特徴パラメタとして、表4、表5に示すものを用いた。ただし、これら特徴パラメタは、著者が主観的に選択したものであり、必ずしも最適なものであるという保証はない。特徴パラメタの中には、ユーザの情報選択の判断に対する意思決定に強く影響を与えるものもあれば、逆にモデル構築に際してノイズとなってしまうものもあると考えられる。こういったノイズとなるような特徴パラメタは、システムの推薦精度を低下させてしまうため、システム設計の際にはこれらの特徴パラメタを最適化する必要がある。本実験では、各特徴パラメタがモデルの識別性能に寄与している度合を分析することによって、特徴パラメタの最適化に関する検討を行う。

4.7.1 実験方法

表4、表5に示した特徴パラメタを対象に分析を行う。分析対象のモデルは、4.6節の実験において構築した被験者9名それぞれのCPモデルとする。

以下、実験手順を示す。

- (1) あらかじめ4.6節の実験で構築したCPモデルの識別性能を得ておき、これを標準時の識別性能とする。
- (2) 表4に示す飲食店特徴パラメタを[ジャンル]、[予算]、[店の特徴]、[雰囲気]の4組に、表5に示すコンテキスト特徴パラメタを[月]、[曜日]、[時刻]、[気象情報]、[ユーザ情報]、[同伴者情報]の6組に、それぞれ分ける。
- (3) 学習データを構成する特徴パラメタから、上記の特徴パラメタを1組ずつ排除する。
- (4) 1組の特徴パラメタが排除された学習データ集合から各被験者のCPモデルを新たに構築し、そのモデルの識別性能を得る。
- (5) 特徴パラメタ排除後の識別性能と標準時の識別性能とを比較する。
- (6) これをすべての組について行う。

以上の実験の結果，排除されることにより識別性能を標準時よりも低下させた特徴パラメタは，もともとモデルの識別性能の向上に寄与していた特徴パラメタであるといえる．一方で，排除されることにより識別性能を標準時よりも向上させた特徴パラメタは，モデル構築に際してノイズとなっていた特徴パラメタであるといえる．このような考えに基づいて特定された，ノイズとなった特徴パラメタを排除していくという方法によって，特徴パラメタの最適化を行う．

4.7.2 実験結果

図 12 に，飲食店特徴パラメタとコンテキスト特徴パラメタを 1 組ずつ排除したときのモデルの識別性能を示す．ここでの識別性能は，未知データに対する識別性能すなわち汎化性能であり，9 名の被験者の CP モデルの識別性能の平均を示している．横軸は排除した特徴パラメタの組を表す．図中の破線は特徴パラメタを排除しないとき，つまり標準時の識別性能を示す．

図 12 から，特徴パラメタを排除することによって，識別性能を標準時よりも低下させたものとして，低下率が大きいものから順に，

1st: [ジャンル]

2nd: [同伴者情報]

3rd: [時刻]，[ユーザ情報]

5th: [雰囲気]

6th: [予算]

の 6 組が特定された．これらは，モデルの識別性能の向上に寄与していた特徴パラメタであるといえる．ここで，この低下率が大きいものほど，識別性能の向上に大きく寄与していたものであり，特に [ジャンル] パラメタは，本実験においては，その向上に大きく寄与していたといえる．

一方，特徴パラメタを排除することによって，識別性能を標準時よりも向上させたものとして，向上率が大きいものから順に，

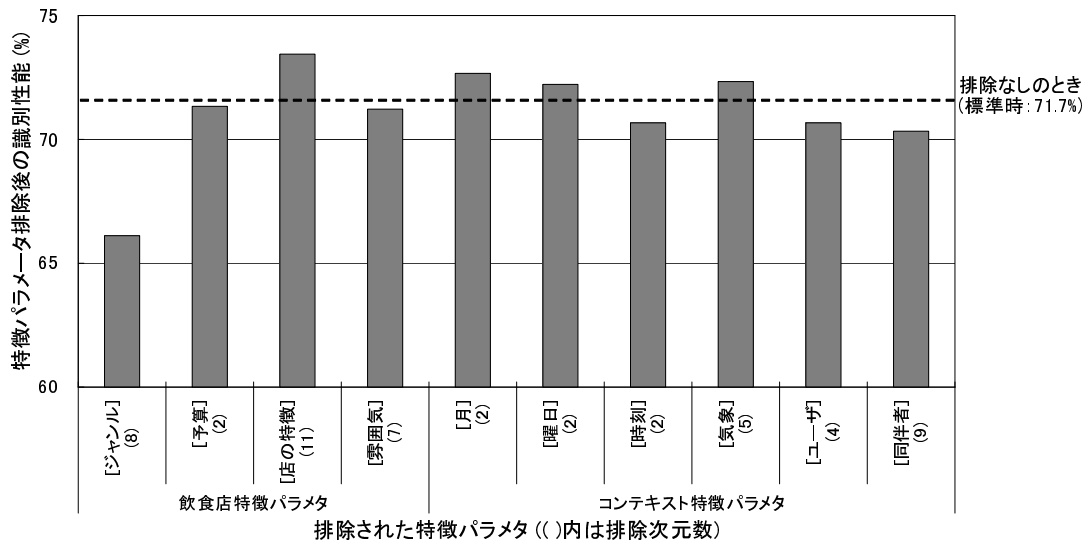


図 12 特徴パラメタがモデルの識別性能に与える影響の分析

1st: [店の特徴]

2nd: [月]

3rd: [気象情報]

4th: [曜日]

の4組が特定された。これらは、モデル構築に際してノイズとなっていた特徴パラメタであるといえる。

さらに、この4組のノイズとなった特徴パラメタと[予算]パラメタとを累積的に排除していったときの識別性能の変化を表6に示す。この結果から、各被験者のCPモデルから、[店の特徴]、[月]、[気象情報]、[曜日]の4組を排除したときに、最大の識別性能(74.7%)が得られた。

4.7.3 考察

以上の実験結果に基づいて、特徴パラメタ最適化に関する考察を行う。

表 6 特徴パラメタを累積的に排除したときの識別性能の変化

排除された特徴パラメタの組 (×:排除パラメタ)					識別性能 (%)
店の特徴	月	気象情報	曜日	予算	
-	-	-	-	-	71.7
×	-	-	-	-	73.4
×	×	-	-	-	73.8
×	×	×	-	-	74.1
×	×	×	×	-	74.7
×	×	×	×	×	73.6

本実験の被験者に対して、「情報選択の判断において、参考にした項目の順位」に関してヒアリング調査を行った。この調査の結果、被験者が意識的に参考にした項目の順位は、飲食店情報の項目に関して、

1st: [ジャンル]

2nd: [予算]

3rd: [雰囲気]

コンテキストの項目に関して、

1st: [同伴者情報]

2nd: [時刻]

3rd: [ユーザ情報]

となり、この結果と本実験結果から得られた順位とがほぼ一致することを確認した。

この結果から、特徴パラメタがモデルの識別性能に与える影響を分析することによって、ユーザの情報選択の判断基準となる特徴パラメタを抽出することが可能であるといえる。本来、人の意思決定過程をモデル化するためには、行動心理など複雑な事象に基づいた分析が求められると考えられるが、本実験結果は、このような複雑な分析を、今回のような簡単な方法により行うことができるということを示唆している。また、表 6 に示すように、ノイズとなった特徴パラメタを

排除していくことによって、識別性能を高めていくことが可能であるという結果が得られたことから、本実験で行った分析方法は、特徴パラメタ最適化手法として、適用し得るものであるといえる。

ただし、本実験で特定した最適パラメタは、あくまでも今回の実験環境および実験条件によるものであるため、この最適パラメタをそのまま実用システムに適用できるとは限らない。システムを実用化するためには、例えば、システム運用前に一定のモニタ期間を設置するなど、より現実に近い環境において、特徴パラメタを最適化する必要がある。

また、扱う特徴パラメタに関しては、各特徴ベクトル次元の直交性が保証されていることが望ましい。本研究で採用した特徴パラメタに関しても、できる限り独立関係になるよう選定したが、主観的な要素も含んでいるため、完全な直交性を保証することは困難である。したがって、今後も、できる限り各パラメタの直交性を確保することを目指すことで、推薦精度の低下を防ぐことが必要であると考えている。

4.8 結言

本章では、コンテキストに応じて変化するユーザの嗜好をモデル化する方法である、コンテキスト依存型ユーザ嗜好モデル化方式およびこれを適用したコンテキスト依存型情報推薦方式を提案した。この情報推薦方式には次の二つの方式がある。

- (a) コンテキスト依存型情報フィルタリング方式 (C-IF)
- (b) コンテキスト依存型協調フィルタリング方式 (C-CF)

本章ではまた、この推薦方式を飲食店情報推薦に対して適用した検証実験を行った。この検証実験より得られた成果と知見を以下に示す。

- 他のクラス分類手法との比較評価の結果、SVMは未学習データに対する汎化性能に優れていることが実証でき、情報推薦方式に対して適用することの妥当性を裏付けた。

- 二つの情報推薦方式, C-IF および C-CF に対して以下の特長を明らかにした。
 - － 既知エリアを推薦対象とした場合には, C-IF による情報推薦が有効に働く。
 - － 未知エリアを推薦対象とした場合には, C-CF による情報推薦が有効に働く。
- 推薦対象のアイテムデータおよびユーザコンテキストの特徴パラメタの最適化に関する検討を行い, 各特徴パラメタが識別性能に与える影響を分析することによって, 特徴パラメタの最適化が可能であるということを示した。

5. コンテキスト依存型ランキング方式

5.1 緒言

情報推薦技術の目的の一つとして、情報探索行為におけるユーザ負担の軽減がある。ユーザの情報探索行為に対する負担を軽減するために、本章では、第4章の推薦方式により選定された推薦候補アイテム集合に対するランキングを作成することにより、ユーザに提示するアイテム数を上位いくつかに絞り込む方法を取る。

推薦候補アイテム集合に対する単純なランキング方法としては、例えば飲食店を推薦対象とした場合には、「料金の安さ」を基準にしたランキング、ユーザの「現在地からの距離の近さ」を基準にしたランキングなどが挙げられる。しかしながら、ユーザの情報選択における価値判断基準はユーザコンテキストに応じて変化するため、アイテムのどの属性を基準にランキングを作成すべきかは、ユーザ個人によっても、そのときのユーザコンテキストによっても変わってくる。あるユーザは、「安さ」重視で店を選ぶかもしれないし、また別のユーザは、「近さ」重視で店を選ぶかもしれない。また、同じユーザであっても、「一人でいるときは安い店」を選ぶかもしれないし、「恋人といるときは高級な店」を選ぶかもしれない。つまり、ユーザがそのとき直面しているコンテキストにおいて、何を重視しているかを考慮したうえで、ランキングを行う必要があると考える。

そこで、本章では、ユーザの価値判断基準がユーザコンテキストによって変化することを考慮に入れた、ユーザの価値判断基準モデルというものを提案する。このモデルに基づいて推薦候補アイテム集合に対するスコア付けを行うことにより、ユーザコンテキストに合ったランキングを提供することが可能となる。本論文では、この価値判断基準モデルに基づいたランキング方式のことを、コンテキスト依存型ランキング方式とよぶ。

コンテキスト依存型ランキング方式の流れは、次のとおりである。

- (1) ユーザの学習データからのコンテキスト依存型価値判断基準モデルを構築する。
- (2) 価値判断基準モデルに基づき、推薦候補アイテム集合をランキングする。

以下、本章の構成を示す。5.2 節では、基本的事項として、既存のランキング手法について説明する。5.3 節では、ユーザの価値判断基準モデルの定義を行い、提案方式であるコンテキスト依存型ランキング方式について説明する。5.4 節では、提案方式に対する評価実験および考察を行い、5.5 節で、本章をまとめる。

5.2 基本的事項

本節では、既存のランキング手法について取り上げる。

単純なランキング手法として、ベクトル空間モデル [5][6][7] やベイズ分類器 [19][20] を用いた方法が挙げられる。

ベクトル空間モデル [5][6][7] は、情報検索における検索モデルとして知られている。検索対象の文書および検索質問を特徴ベクトルで表現し、ベクトル間の類似度を計算することにより検索質問に適した文書を提示する。この類似度の高さに基づき、検索質問に合ったランキングを行うことが可能である。

ここで、ユーザ嗜好を何らかの方法により特徴ベクトル化することによって、ユーザ嗜好に合ったランキング結果を提示することができると考えられる。しかしながら、ユーザの嗜好はユーザコンテキストにより多様に変化するため、ユーザ嗜好を一つの特徴ベクトルで表現することは容易でない。ユーザ嗜好が適切に表現できなければ、これに伴うランキング結果に対する信頼性も乏しい。

ベイズ分類器 [19][20] は、確率モデルに基づく分類手法であり、対象となるデータを学習データに基づいて確率的に分類するものである。スパムメールのフィルタリングなどに応用されている例が挙げられる [20]。ここでも、ユーザ嗜好を確率モデルにより表現することにより、ユーザ嗜好に合わせたランキングを提供することが可能となるが、先と同様、ユーザ嗜好をより適切に表現するためには、ユーザ嗜好のコンテキスト依存性を考慮に入れる必要がある。

第4章で述べたように、本論文では、ユーザ嗜好のモデル化方式としてSVM[18]を適用している。SVMを用いたランキング手法としては、文書検索の分野において、いくつかの研究が行われている。

平尾ら [72][73] は、重要文抽出に SVM を用いている。重要文抽出を、文書中の各文に対し、重要 (正事例)、非重要 (負事例) のクラスを付与する 2 クラス分

類問題として扱っている。この研究では、SVMの識別関数（識別超平面）による値に基づき、文の重要度の算出を行っている。具体的には、識別関数による値が大きい（すなわち特徴空間上で識別超平面から遠くにある）事例ほど、重要であるとみなしている。しかし、SVMの識別関数による値をそのままランキングに適用させることが妥当であるかどうかについては、理論的な根拠は明らかでない[74][75]。

Westonら[8][9]は、多クラス分類問題に対応したMulti-Class SVMを提案している。この手法では、 k クラス分類問題を k 段階の2クラス分類問題として扱っている。例えば、第 n 段階の分類問題は、クラス n とそれ以外の $k-1$ 個のクラスとの2クラス分類問題とみなすことができる。ここで、クラスをランクとして扱うことにより、Multi-Class SVMをランキング手法として適用することが可能である。

Herbrichら[76][77][78]は、Ranking SVMを提案している。この手法では、事例のペアを学習データとして用い、SVMによりこのペアの順序関係を学習することにより、新しい事例に対するランク付けを行う。

Multi-Class SVMやRanking SVMは、あらかじめ学習データとして与えられるアイテム間に順序関係があることを前提とした手法である。しかしながら、本論文で扱う情報推薦という問題においては、ユーザの情報選択における価値判断基準は、ユーザがそのときおかれているコンテキストによって変化するため、コンテキストが変われば、アイテム間の順序関係も変わってしまう。したがって、Multi-Class SVMやRanking SVMをそのまま情報推薦におけるランキング手法として適用することは妥当でない。

これに対し、本章での提案方式は、ユーザコンテキストによってユーザの価値判断基準が変化することを考慮しているため、推薦時のユーザコンテキストに相応しいランキングを提供することが可能である。

5.3 コンテキスト依存型ランキング方式

本節では、まずユーザのコンテキスト依存型価値判断基準モデルを提案する。そして、このモデルに基づくコンテキスト依存型ランキング方式について説明

する。

本節の構成は以下のとおりである。

- (1) ユーザのコンテキスト依存型価値判断基準モデルの定義。
- (2) ユーザの学習データからの価値判断基準モデルの構築方法の説明。
- (3) コンテキスト依存型価値判断基準モデルに基づくランキング方式の説明。

まず、5.3.1 項において、ユーザのコンテキスト依存型価値判断基準モデルの定義を行う。つづいて、5.3.2 項において、このコンテキスト依存型価値判断基準モデルをユーザの学習データから構築する方法について述べる。5.3.3 項において、コンテキスト依存型価値判断基準モデルに基づくランキング方式について説明する。

5.3.1 ユーザのコンテキスト依存型価値判断基準モデル

ユーザがあるアイテム集合の中から欲しいアイテムを選択する際、そのユーザは何らかの独自の判断基準を持っているといえる。例えば、数多くある飲食店の中から行きたい店を選ぶ際、ユーザは飲食店に対して、「安い店が良い」、「近い店が良い」、「バーが良い」といった判断基準を持つ。このユーザが持っている判断基準のことを、本章では、ユーザのアイテム選択に対する価値判断基準とよぶ。

図 13 は、あるユーザの飲食店選択に対する価値判断基準を図で表したものである。図 13 (a) は、ユーザ A、ユーザ B およびユーザ C の 3 人のユーザが存在し、それぞれ「安い店が良い」、「安くて近い店が良い」、「5,000 円程度のバーが良い」といった価値判断基準を持っていることを示している。

それぞれのユーザの価値判断基準をモデル化したものが図 13 (b) である。このモデルの横軸は各アイテムパラメタ (ここでは、[料金], [現在地からの距離], [料理ジャンル]) のとる値を示し、縦軸はそのパラメタ値に対するユーザの満足度 (正 ⇒ 満足, 負 ⇒ 不満足) を示している。このようにユーザの価値判断基準を表現したモデルを、本論文では、ユーザの価値判断基準モデルとよぶ。このモデルは、ユーザ A のように飲食店の唯一つのパラメタ (ここでは [料金]) により表現され

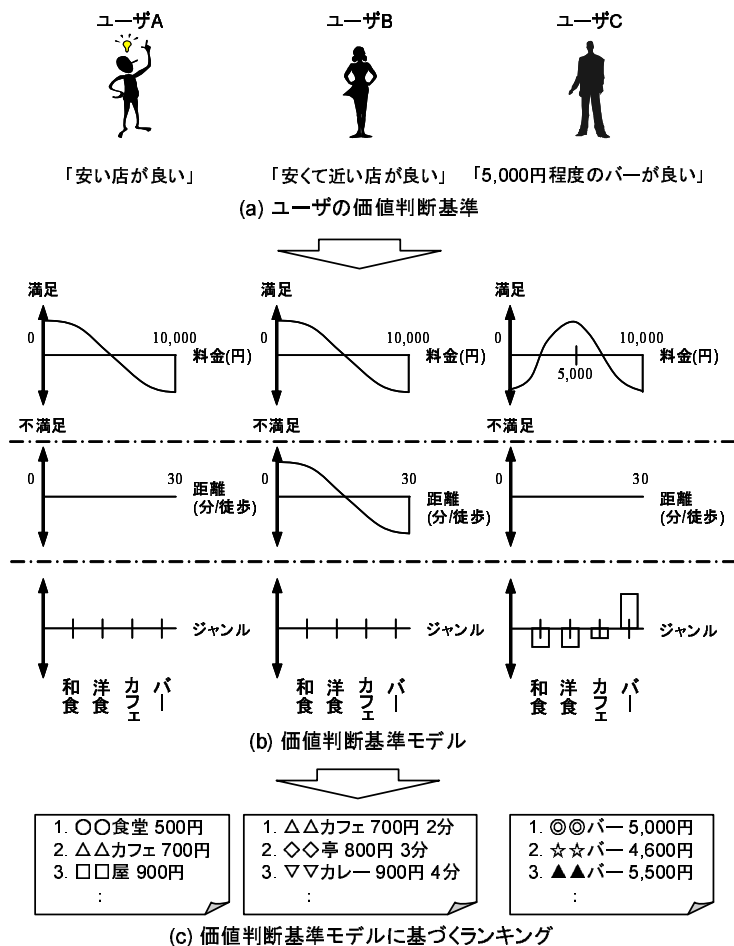


図 13 ユーザのアイテム選択に対する価値判断基準 (飲食店選択の例)

る場合もあれば、ユーザ B やユーザ C のように複数のパラメタにより複合的に表現される場合もある。

図 13 (c) は、各ユーザの価値判断基準モデルに沿って推薦候補である飲食店集合に対しランキングを行っている例である。このようにユーザの価値判断基準を適切にモデル化することができれば、価値判断基準に沿ったランキングを行うことは容易である。

しかしながら、ユーザの価値判断基準はユーザコンテキストに応じて変化するものである。例えば、同じユーザ A であっても、一人では「安い店が良い」、友人といるときには「近い店が良い」、恋人といるときには「5,000 円程

度のバーが良い」といったようにユーザコンテキストによって価値判断基準は大きく変化するものと考えられる。したがって、推薦候補アイテム集合に対するランキングも必然的にユーザコンテキストに応じて変えなければならない。このことから、ユーザに対し、一意な価値判断基準モデルを作ることは適切でなく、推薦時のユーザコンテキストに合ったモデルを作る必要がある。

提案方式では、ユーザコンテキストに適切なランキングを提示するために、推薦時のユーザコンテキストに依存したユーザの価値判断基準モデルを構築する。まず、5.3.2項では、ユーザの学習データからユーザコンテキストに依存した価値判断基準モデルの構築方法について述べる。つづいて、5.3.3項では、この価値判断基準モデルに基づいたランキング方法について説明する。

5.3.2 ユーザの学習データからの価値判断基準モデルの構築

本項では、ユーザの学習データからユーザコンテキストに依存した価値判断基準モデルを構築する方法について述べる。本項の流れは次のとおりである。

- (1) ユーザの学習データ形式の定義
- (2) 当該コンテキストに関連する学習データの特定方法の説明
- (3) 当該コンテキストにおけるユーザの価値判断基準モデルの構築方法の説明

以下、各項目について述べる。

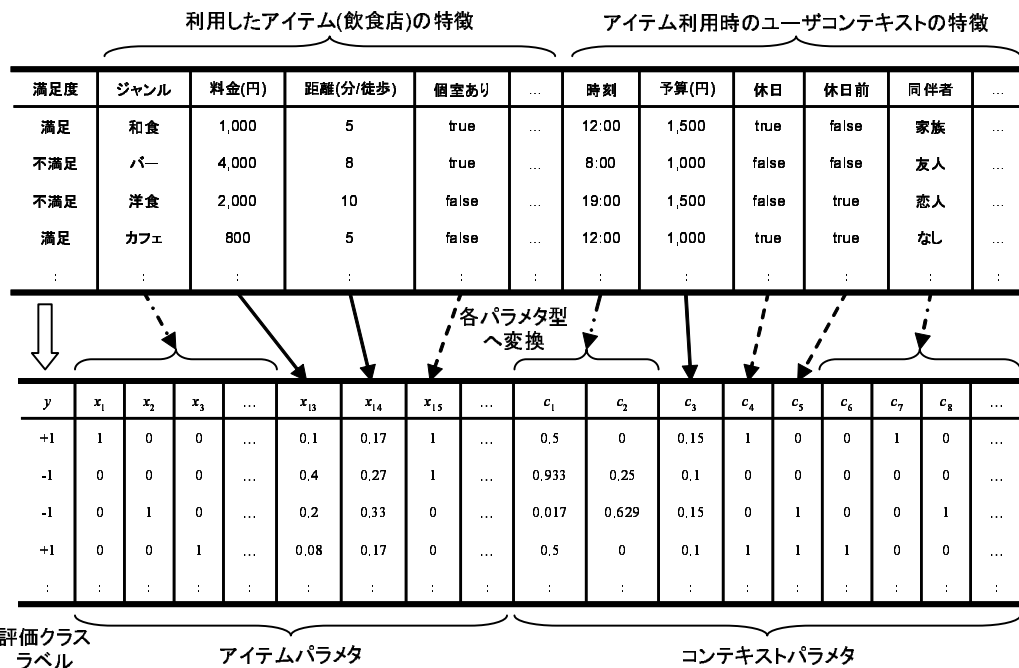
ユーザの学習データ形式

まず、価値判断基準モデルの構築方法について説明する前に、元となるユーザの学習データの形式について説明する。

図 14 は、推薦対象アイテムを飲食店とした場合の学習データの例である。図 14(a) は、ユーザのアイテム利用履歴データの例を表したものであり、以下の情報を持つ。

- あるユーザコンテキストにおいて利用したアイテムに対する評価 (“満足”・“不満足”)

(a) ユーザのアイテム利用履歴データの形式



(b) ユーザの学習データの形式

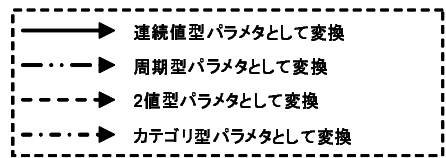


図 14 ユーザのアイテム利用履歴データと提案方式で扱う学習データ形式への変換

- 利用したアイテムの特徴に関する情報
- アイテム利用時のユーザコンテキストの特徴に関する情報

この利用履歴データを提案方式で扱える形にするために，図 14(b) に示すような学習データ形式に変換する．学習データは次のような情報から構成される．

- $y = \{+1, -1\}$: 利用したアイテムに対する評価クラスラベル
- $x_1, x_2, \dots, x_p (0 \leq x_i \leq 1)$: 利用したアイテムの特徴を表現する p 次元特徴パラメタ (つまり，アイテムパラメタ)

- $c_1, c_2, \dots, c_q (0 \leq c_j \leq 1)$: アイテム利用時のユーザコンテキストを表現する q 次元特徴パラメタ (つまり, コンテキストパラメタ)

図 14 に示すように, アイテムパラメタおよびコンテキストパラメタは, その表現方法により, 以下のいずれかの型のパラメタとして扱われる.

- 連続値型パラメタ : $[0, 1]$ の連続値により表現 ([料金] や [予算] など).
- 周期型パラメタ : $([0, 1], [0, 1])$ の 2 次元の連続値型パラメタにより表現 ([時刻] など)
- 2 値型パラメタ : $\{0, 1\}$ の 2 値により表現 ([個室あり] や [休日] など).
- カテゴリ型パラメタ : 3 次元以上の子パラメタを持ち, 該当する子パラメタに “1” を, それ以外の子パラメタに “0” を与えることにより表現 ([料理ジャンル] や [同伴者] など)

ここで, 周期型パラメタは, [時刻] や [月], [曜日] など周期性を持つものを表現するために扱われる. 例えば, [時刻] は “23 : 59” の 1 分後が “0 : 00” となるように, 始点と終点が一致するような周期性を持っている. この周期性を表現するために, 図 15 に示すように [時刻] を二つの次元 (時刻 1, 時刻 2) により表現している. 例えば, “3 : 00” は $(0.854, 0.854)$ で表現し, “6 : 00” は $(1, 0.5)$ で表現する. また, “24 : 00” = “0 : 00” は $(0.5, 1)$ で表現することができ, 始点と終点を同値として扱うことができる.

また, カテゴリ型パラメタの例としては [同伴者] が挙げられる. [同伴者] は親パラメタであり, 子パラメタとして {[なし], [家族], [恋人], [友人], ...} を持つ. 家族といることを表現するためには, 子パラメタ [家族] に対し “1” を, それ以外の子パラメタに対し “0” を与える.

本章では, 以降, カテゴリ型パラメタは, [同伴者:家族] のように, [親パラメタ名:子パラメタ名] と表記することにする. また, パラメタの持つ値を示すときは, [予算]=“1,500 円”, [休日]=“1”, [同伴者:家族]=“1” のように, [パラメタ名]=“パラメタ値” と表記する.

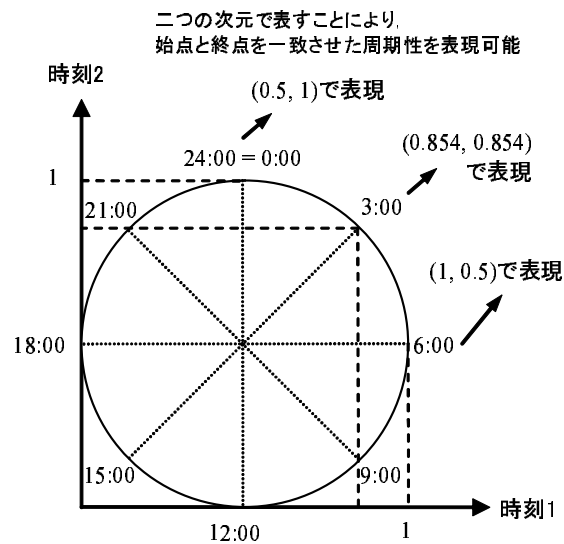


図 15 周期型パラメタの表現 ([時刻] の例)

当該コンテキストに関連する学習データの特定

ユーザコンテキストに依存した価値判断基準モデルを構築するための前処理として、当該コンテキストに関連する学習データの特定を行う必要がある。

当該コンテキストに関連する学習データを特定する方法として、単純には、当該コンテキストと完全一致するような学習データを選べばよい。例えば、いま、ユーザ A の推薦時のユーザコンテキストが $C_A = \{[\text{同伴者:恋人}] = "1", [\text{休日}] = "1", [\text{予算}] = "5000" \dots\}$ であった場合、ユーザの全学習データの中から、 C_A と一致するユーザコンテキストを持つ学習データを選ぶ。

しかし、すべてのコンテキストパラメタの値が一致するような学習データの数は、ユーザの全学習データ数に対して極少であると考えられる。極少な学習データからでは、適切な価値判断基準モデルを構築することは困難である。したがって、ここでは、すべてのコンテキストパラメタの値の一致をみるのではなく、着目すべきコンテキストパラメタを選定したうえで、このパラメタの値の一致をみる方法を考える。このように着目すべきコンテキストパラメタの数を絞り込むことによって、確保できる学習データ数が増加する。

そこで、本章では、多次元の特徴ベクトルの中から重要な属性を選択する方法

(特徴選択手法 [79]) の一つとして、一般によく用いられる情報利得 [23] を用いて、着目すべきコンテキストパラメタの選定を行う。各特徴パラメタに対し、情報利得を求め、最も大きい情報利得を持つ特徴パラメタを選定する。

いま、推薦時のユーザコンテキストが $C = \{[\text{同伴者:恋人}] = \text{“1”}, [\text{休日}] = \text{“1”}, [\text{予算}] = \text{“5000”} \dots\}$ であったとする。このとき、各コンテキストパラメタ [同伴者:恋人]、[休日]、[予算] について、情報利得を求める。

以下、パラメタの型ごとに、情報利得の求め方について説明する。

< カテゴリ型・2 値型パラメタの情報利得の算出方法 >

カテゴリ型パラメタである [同伴者:恋人] を c_i とすると、このパラメタの情報利得 $Gain_i$ は次式により求められる。

$$Gain_i = E_1 - E_{2i} \quad (4)$$

$$E_1 = -\frac{n^+}{N} \log_2 \frac{n^+}{N} - \frac{n^-}{N} \log_2 \frac{n^-}{N} \quad (5)$$

$$\begin{aligned} E_{2i} = & -\frac{n|_{c_i=1}}{N} \frac{n^+|_{c_i=1}}{n|_{c_i=1}} \log_2 \frac{n^+|_{c_i=1}}{n|_{c_i=1}} \\ & -\frac{n|_{c_i=1}}{N} \frac{n^-|_{c_i=1}}{n|_{c_i=1}} \log_2 \frac{n^-|_{c_i=1}}{n|_{c_i=1}} \\ & -\frac{n|_{c_i=0}}{N} \frac{n^+|_{c_i=0}}{n|_{c_i=0}} \log_2 \frac{n^+|_{c_i=0}}{n|_{c_i=0}} \\ & -\frac{n|_{c_i=0}}{N} \frac{n^-|_{c_i=0}}{n|_{c_i=0}} \log_2 \frac{n^-|_{c_i=0}}{n|_{c_i=0}} \end{aligned} \quad (6)$$

ここで、 N は全学習データ数、 n^+ 、 n^- は正クラス、負クラスの学習データ数、 $n|_{c_i=v}$ はコンテキストパラメタ $c_i = v (v = \{0, 1\})$ の条件下における学習データ数、 $n^+|_{c_i=v}$ 、 $n^-|_{c_i=v}$ は $c_i = v$ の条件下における正クラス、負クラスの学習データ数を表す。

なお、2 値型パラメタである [休日] についても、カテゴリ型パラメタと同様に情報利得を求めることができる

< 連続値型・周期型パラメタの情報利得の算出方法 >

連続値型パラメタである [予算] を c_j とすると、上記の式 (6) は、次式のように

なる．

$$\begin{aligned}
 E_{2i} = & -\frac{n|_{c_j \approx w}}{N} \frac{n^+|_{c_j \approx w}}{n|_{c_j \approx w}} \log_2 \frac{n^+|_{c_j \approx w}}{n|_{c_j \approx w}} \\
 & -\frac{n|_{c_j \approx w}}{N} \frac{n^-|_{c_j \approx w}}{n|_{c_j \approx w}} \log_2 \frac{n^-|_{c_j \approx w}}{n|_{c_j \approx w}} \\
 & -\frac{n|_{c_j \neq w}}{N} \frac{n^+|_{c_j \neq w}}{n|_{c_j \neq w}} \log_2 \frac{n^+|_{c_j \neq w}}{n|_{c_j \neq w}} \\
 & -\frac{n|_{c_j \neq w}}{N} \frac{n^-|_{c_j \neq w}}{n|_{c_j \neq w}} \log_2 \frac{n^-|_{c_j \neq w}}{n|_{c_j \neq w}}
 \end{aligned} \tag{7}$$

ここで， w は，現在のユーザコンテキストのパラメタ c_j の値である．この例では，[予算]=“5000 円” であるため，正規化し $w = 0.5$ となる．

なお，条件としている $c_j \approx w$ については，その許容範囲を設け，次のように定義する．

$$0.5 - \alpha \leq c_j \leq 0.5 + \alpha (0 \leq \alpha \leq 1) \tag{8}$$

ここで， α の設定方法については以下のとおりである．

α の設定方法

α を小さくすると，より現在のユーザコンテキストに限定した学習データを選定することができるため，そのコンテキストによるユーザの価値判断への影響をより強くモデルに反映させることができる．一方で，コンテキストを限定することにより確保できる学習データ数が少なくなり，コンテキスト依存の価値判断基準モデルの構築が困難になる．したがって，モデル構築に十分な学習データ数を確保しつつ，コンテキストによる影響をより強く反映させられるような α を求める必要がある．

α の設定目標としては，モデル構築に十分な学習データ数を確保できる範囲で，できるだけ小さい値を選択する．具体的には，次の手順により α を設定する．

- (1) モデル構築に必要となる最小限の学習データ数（後述）を決める．
- (2) 現在のコンテキストパラメタの値を基準に， α を 0.1 刻み等，徐々に大きくしていきながら，そのとき確保できる学習データ数をみる．

(3) step 1 で決めた最小限の学習データ数を確保できる最小の α を採用する .

モデル構築に必要な最小限の学習データ数は、推薦対象とするコンテンツ等にも依存することから、情報推薦システムの設計時において実証実験を通じて求める必要がある . 具体的には、学習データ数 $\{10, 20, 30, \dots\}$ とテストデータに対する識別性能の関係から、識別性能が収束する最小の学習データ数を選ぶ . なお、5.4 節における評価実験においては、予備実験により最小限の学習データ数を求めている .

ここでは、連続値型パラメタについての情報利得の算出方法を示したが、周期型パラメタに関しても、二つの次元を同時に扱うことを除き、連続値型パラメタと同様の方法で情報利得を求めることができる .

以上の方法により、各コンテキストパラメタ c_i の情報利得 $Gain_i$ を算出し、最も大きい情報利得 $Gain^*$ を持つコンテキストパラメタ c^* を、着目するコンテキストパラメタとして選定する . そして、全学習データの中から、着目するコンテキストパラメタの値が推薦時のユーザコンテキストのパラメタの値と一致する学習データを特定する .

例えば、今回の例の場合、[同伴者] が着目するコンテキストパラメタとして選定された場合、[同伴者]=“1”の条件を満たす学習データを選定する . [予算] が着目するコンテキストパラメタとして選定された場合、 $0.5 - \alpha \leq [\text{予算}] \leq 0.5 + \alpha$ ($0 \leq \alpha \leq 1$) (正規化後) の条件を満たす学習データを特定する .

当該コンテキストにおけるユーザの価値判断基準モデルの構築

特定された当該コンテキストに関連する学習データに基づいて、当該コンテキストにおけるユーザの価値判断基準モデルを構築する .

ユーザの価値判断基準モデルは、図 16 に示すように、着目しているアイテムパラメタ軸上における正負データ (ユーザが“満足”・“不満足”と評価付けしたデータ) の分布に基づいて構築される . 図 16(a) は、例えばアイテムパラメタ [料金] の軸上における正負データの分布を示したものである . この例では、正データが [料金]=5,000 円あたりに集中しているため、ユーザは 5,000 円程度の店に対し満足しているということが読み取れる .

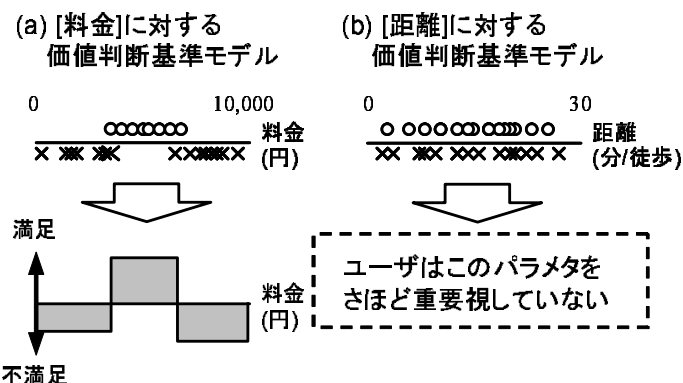


図 16 アイテムパラメータ軸上における正負データの分布に基づく価値判断基準モデルの構築

一方で、着目するアイテムパラメータ軸によっては、図 16(b) のように、正データおよび負データが混在してしまい、有効な価値判断基準モデルを構築できない場合もありうる。しかし、これはそもそもこのアイテムパラメータからでは正負クラスの判別が不可能であるということの意味し、ユーザはこのアイテムパラメータ(図 16(b) の例では [距離]) をさほど重要視していないと捉えることができる。したがって、このようにユーザが重要視していないアイテムパラメータに対しユーザの価値判断基準モデルを構築することはあまり意味がない。

そこで、提案手法では、ユーザが重要視しているアイテムパラメータを選定し、このアイテムパラメータを対象にユーザの価値判断基準モデルを構築する。以下、まず、ユーザが重要視しているアイテムパラメータの選定方法について述べ、つづいて、このアイテムパラメータを対象としたユーザの価値判断基準モデルの構築方法について説明する。

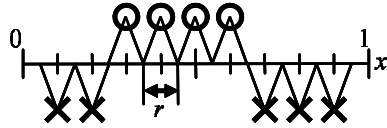
ユーザが重要視しているアイテムパラメータの選定

あるアイテムパラメータに対しユーザがどの程度重要視しているかを示す指標を、このアイテムパラメータに対する重要度と定義する。図 16(a) のように、着目しているアイテムパラメータ軸上で、正および負の学習データが明確に分離されているような場合には、この重要度は大きくなる。なぜなら、このパラメータによって、

$$n^+ = 4, n^- = 5$$

When $r = 0.1$

$$m_i^+ = 3, m_i^- = 3, m_i^* = 2$$



アイテムパラメタ x_i の重要度

$$w_{xi} = \left(\frac{2 \times 3}{4(4-1)} + \frac{2 \times 3}{5(5-1)} - \frac{2 \times 2}{4 \times 5} \right) = +0.6$$

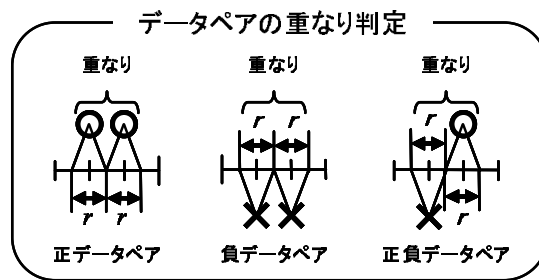


図 17 アイテムパラメタの重要度の算出例とデータペアの重なり判定 (連続値型パラメタの場合)

正および負のデータを分離することが容易になるからである。一方，図 16(b) のように，正データおよび負データが混在しているデータが多い場合には，この重要度は小さくなる。

このことを踏まえ，アイテムパラメタ x_i に対する重要度 w_{xi} を，次式のように定義する。

$$w_{xi} = \begin{aligned} & \text{正データの密集度合い} \\ & + \text{負データの密集度合い} \\ & - \text{正負データの混在度合い} \end{aligned} \quad (9)$$

図 17 は，式 (9) にしたがって，アイテムパラメタ x_i の重要度を算出している例である。式 (9) の第 1 項 (第 2 項) は，アイテムパラメタ x_i 軸上で，正 (負) データがどの程度密集しているかを，また，第 3 項は，正負データがどの程度混在し

ているかを示し，具体的には，次式で求められる．

$$\text{正データの密集度合い} = \frac{2m_i^+}{n^+(n^+ - 1)} \quad (10)$$

$$\text{負データの密集度合い} = \frac{2m_i^-}{n^-(n^- - 1)} \quad (11)$$

$$\text{正負データの混在度合い} = \frac{2m_i^*}{n^+n^-} \quad (12)$$

ここで， n^+, n^- はそれぞれ正データ数，負データ数である．図 17 の例では，それぞれ $n^+ = 4, n^- = 5$ となる．正データおよび負データの密集度合いを算出するために，図 17 に示すように任意のデータ幅 $r (0 \leq r \leq 1)$ (後述) を設ける．この幅 r を考慮したとき，重なり合っているデータのペアの数を数える．ここでは，重なり合っているデータのペアをデータペアとよび，図 17 の例に示すように，

- 正データペア：正データ同士が重なり合っているデータペア
- 負データペア：負データ同士が重なり合っているデータペア
- 正負データペア：正負データが重なり合っているデータペア

とよぶことにする．式 (10), (11), (12) の m_i^+, m_i^-, m_i^* は，それぞれ正データペア，負データペア，正負データペアの数である．図 17 の例では，それぞれ $m_i^+ = 3, m_i^- = 3, m_i^* = 2$ となる．

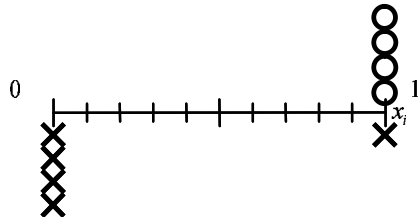
図 16(a) のように，ある領域に正 (負) データが密集していれば，第 1 項 (第 2 項) の値は大きくなり，図 16(b) のように，ある領域に正負データが混在していれば，第 3 項の値が大きくなる．したがって，(第 1 項+第 2 項) の値と第 3 項の値との差が大きいほど，着目しているアイテムパラメタは重要であるということを意味する．

なお，図 17 は連続値型パラメタでの算出例を示した．2 値型パラメタおよびカテゴリ型パラメタの場合は，パラメタ値のとりうる値が $\{0, 1\}$ のどちらかであるため，重要度の算出は連続値型パラメタの場合よりも単純になる．図 18 は，ある二つの 2 値型のアイテムパラメタの重要度算出例を示したものである．図 18 左の例のように，正負データの分布が偏っている方が，重要度が高くなっていることが確認できる．

$$n^+ = 4, n^- = 5$$

When $r = 0$

$$m_i^+ = 6, m_i^- = 6, m_i^* = 4$$



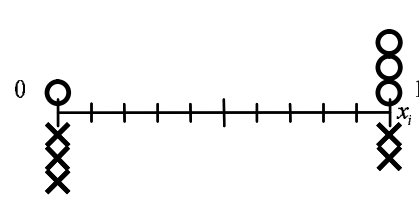
アイテムパラメタ x_i の重要度

$$w_{xi} = \left(\frac{2 \times 6}{4(4-1)} + \frac{2 \times 6}{5(5-1)} - \frac{2 \times 4}{4 \times 5} \right) = +1.2$$

$$n^+ = 4, n^- = 5$$

When $r = 0$

$$m_i^+ = 3, m_i^- = 4, m_i^* = 9$$



アイテムパラメタ x_i の重要度

$$w_{xi} = \left(\frac{2 \times 3}{4(4-1)} + \frac{2 \times 4}{5(5-1)} - \frac{2 \times 9}{4 \times 5} \right) = 0$$

図 18 アイテムパラメタの重要度の算出例 (2 値型およびカテゴリ型の場合)

この重要度が高い一つ以上のアイテムパラメタを対象に，ユーザの価値判断基準モデルを構築する．

r の設定方法

正データ同士，負データ同士が密集しているようなアイテムパラメタの場合， r が小さいうちは，正データの密集度合い，負データの密集度合いがそれぞれ大きく，正負データの混在度合いが小さくなる． r を徐々に大きくしていくと，重なり合うデータペアも増えるため，その分，正データの密集度合い，負データの密集度合いも大きくなるが，それ以上に正負データの混在度合いも大きくなる．したがって，十分に重なり判定ができる範囲で r を小さくする必要がある．

あるアイテムパラメタ x_i に着目し， r を $0 \leq r \leq 1$ の範囲で，0.01 刻みなどに变化させたとき， w_{xi} が最大となる r を求める． w_{xi} が最大となる r が複数存在する場合は，その中で最小の r を採用する．このときの r を r^* とする． r^* のときの w_{xi} を w_{xi}^* とし，これを着目しているアイテムパラメタに対する重要度とする．

ユーザの価値判断基準モデルの構築

図 19 は，ユーザの価値判断基準モデルの構築過程を示したものである．ここでは，ユーザが重要視しているアイテムパラメタとして， x_i, x_j が判定されたものとする．なお， x_i は連続値型パラメタ， x_j は 2 値型パラメタとする．ユーザは

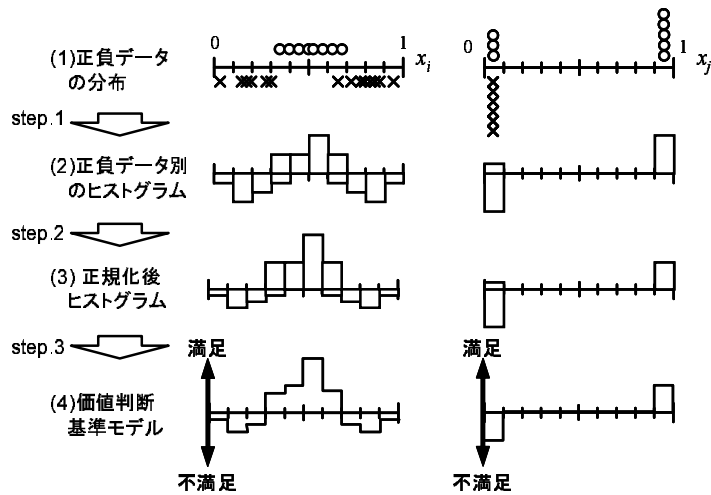


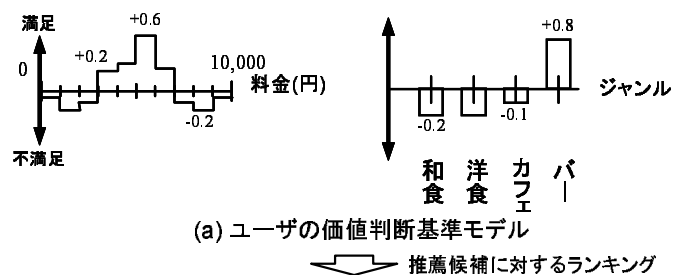
図 19 正負データの分布に基づく価値判断基準モデルの構築手順

重要視しているアイテムパラメタ軸上で、正データが密集している領域において、より満足し、負データが密集している領域において、より不満に感じると考えられる。この観点から正データおよび負データのヒストグラムを作成し、このヒストグラムに基づいて価値判断基準モデルを構築する。

以下、モデルの構築手順を示す。

- step 1 アイテムパラメタ x_i, x_j 軸上の正負データの分布に基づき、正データ、負データ別にヒストグラムをプロットする（ここでは、級の間隔を 0.1 とする）。正データのヒストグラムは正方向に、負データのヒストグラムは負方向にそれぞれプロットする。
- step 2 正データ、負データのそれぞれで、ヒストグラムの総面積が 1 になるように正規化を行う。
- step 3 正データ、負データのヒストグラムを足し合わせるにより、ユーザのアイテムパラメタ x_i, x_j に対する価値判断基準モデルを構築する。

この価値判断基準モデルは、横軸に着目しているアイテムパラメタのとり値を示し、縦軸は、このパラメタの値に依存したユーザの満足度を示す。正方向（満



(a) ユーザの価値判断基準モデル

推薦候補に対するランキング

(b) 推薦候補アイテム集合 (()内は対応するパラメタ値に対するスコア)

店舗名	ジャンル	料金 (円)	距離 (分/徒歩)	...	ランキングスコア計
●●屋	和食 (-0.2)	3,000 (+0.2)	5 (0)	...	0
◎◎バー	バー (+0.8)	5,000 (+0.6)	10 (0)	...	+1.4
△△カフェ	カフェ (-0.1)	700 (-0.1)	2 (0)	...	-0.2
☆☆バー	バー (+0.8)	8,000 (-0.2)	4 (0)	...	+0.6

図 20 推薦候補アイテムに対するランキングスコアの算出

足側) に大きいほど、ユーザはそのアイテムパラメタの値に満足していることを意味する。こうして、当該コンテキストに基づいて構築された価値判断基準モデルを、コンテキスト依存型価値判断基準モデルとよぶ。次項では、この価値判断基準モデルに基づくランキングスコアの算出方法について具体的に説明する。

5.3.3 コンテキスト依存型価値判断基準モデルに基づくランキング方式

ユーザの価値判断基準モデルに基づいて、推薦候補アイテム集合に対し、ランキングスコアを算出する。

ユーザの価値判断基準モデルが図 20(a) に示すようであったとし、また、図 20(b) に示すような推薦候補アイテム集合が与えられたとする。このとき、推薦候補アイテム集合に対するランキングスコアは、図 20 に示すように算出される。

例えば、[ジャンル] が「バー」であるようなアイテムには、[ジャンル] パラメタに対するユーザの価値判断基準モデルから、ランキングスコア “+0.8” が加算される。[料金] に関して、[料金] パラメタに対する価値判断基準モデルから、料金に応じてランキングスコアが加算される。ただし、[距離] パラメタに関しては

ランキングスコアは加算されない。これは、このユーザが [距離] に対しては重要視していないためである。このようにユーザが重要視しているアイテムパラメタのみをランキングスコア加算の対象とすることにより、ユーザの価値判断基準を適切に反映させたランキングスコアの算出が可能となる。

最終的に、こうして得られたランキングスコアを集計することにより、推薦候補アイテム各々のランキングスコアが算出される。このランキングスコアが高いものを上位にし、ユーザに提示する。

5.4 評価実験

提案方式の有効性を評価するために被験者実験を行った。本実験では、飲食店データを推薦対象アイテムとした学習データセットを用いて、被験者の飲食店選択に対する価値判断基準モデルを構築した。そして、学習データセットとは別に作成した評価用データセットに対し、このモデルに基づいたランキングを行い、その有効性を評価した。

5.4.1 実験条件

本実験では、一般向けグルメ情報検索サイトである『ぐるなび』[80] が提供している『ぐるなび API』[81] を利用し、飲食店データセットを作成した。対象エリアとして、さまざまなジャンルの飲食店が豊富に存在する「大阪：なんば」エリアを選択した。このエリアに登録されている全飲食店データの中から 1,000 件を無作為に選んだ。

なお、表 7 は、本実験で用いる学習データセットの形式であり、飲食店データの特徴は、表 7 に示す 14 次元のアイテムパラメタ、ユーザコンテキストは、8 次元のコンテキストパラメタにより、それぞれ表現している。これらのパラメタは、4.7 節で行った被験者実験により、被験者の価値判断基準に影響を及ぼしていると判定されたパラメタを扱っている。

本実験における被験者は、22 歳から 36 歳の男女 10 名である。いずれの被験者も「大阪：なんば」エリアには何度か訪れたことがあり、これらのエリアの雰囲気

表 7 実験で用いた学習データ形式

クラス / パラメタ種別	記号	パラメタ名	パラメタの型	実験で取りうる値		
クラス	class	ユーザ満足度		[5, 4, 3, 2, 1]※		
アイテムパラメタ	x_1	ジャンル:和食	カテゴリ型	いずれか一つ:1 それ以外:0		
	x_2	ジャンル:洋食				
	x_3	ジャンル:中華				
	x_4	ジャンル:アジア・焼肉・エスニック・無国籍				
	x_5	ジャンル:カフェ				
	x_6	ジャンル:和風居酒屋				
	x_7	ジャンル:洋風居酒屋				
	x_8	ジャンル:アジア・無国籍居酒屋				
	x_9	ジャンル:ダイニングバー・ビアレストラン				
	x_{10}	ジャンル:バー・パブ				
	x_{11}	ジャンル:スペース・エンターテイメント				
	x_{12}	ジャンル:ファミレス・ファーストフード				
	x_{13}	料金			連続値型	[0, 1]
	x_{14}	距離			連続値型	[0, 1]
コンテキストパラメタ	c_1	時刻	連続値型	[0.333, 0.500, 0.792]		
	c_2	休日	2値型	[0, 1]		
	c_3	休日前	2値型	[0, 1]		
	c_4	予算	連続値型	[0.1, 0.2, 0.3, 0.5]		
	c_5	同伴者:なし	カテゴリ型	いずれか一つ:1 それ以外:0		
	c_6	同伴者:家族				
	c_7	同伴者:恋人				
	c_8	同伴者:友人				

※比較実験において、提案方式および識別平面からの距離に基づく方法では、ユーザ満足度[5, 4, 3, 2, 1]のうち、3以上を満足(正)クラス、2以下を不満足(負)クラスとして扱う。Multi-Class SVMを用いた方法では、[5, 4, 3, 2, 1]をそのまま5クラスとして扱う。

気などについてはある程度把握している。このように被験者が把握しているエリアを実験対象として選択したのは、研究室内で仮想的に実験を行うため、本実験を行うに当たって、できるだけ被験者にユーザコンテキストを想像しやすくするためである。

5.4.2 実験方法

実験方法について説明する。本実験は以下の4段階のフェーズにしたがって遂行する。

フェーズ1 被験者の学習データセットの作成

フェーズ2 被験者の評価用データセットの作成

フェーズ3 評価用データセットに対するランキング処理

フェーズ4 ランキング結果に対する評価

以下、各フェーズについて詳細に説明する。なお、ここでは、実際に実験の被験者となった被験者 A による実験過程およびデータを例に示しながら説明を行う。

フェーズ1：被験者の学習データセットの作成

まず、以下の手順にしたがって、あらかじめ各被験者の学習データセットを作成する。

- (1) 被験者には、各自「大阪：なんば」エリアにおいて、現実には起こりうるコンテキストを想像してもらい、表7に示すコンテキストパラメタの取りうる値の組合せから、20個のコンテキストパターン $\{C_1, C_2, \dots, C_{20}\}$ を設定してもらおう。この20個のコンテキストパターンは、可能な限り、多様なものになるようにする。
- (2) 被験者には、一つのコンテキストパターンあたり、飲食店データセットの中から無作為に抽出された10件の飲食店データに対し、5段階の満足度評価 $\{5, 4, 3, 2, 1\}$ (5:最も満足, 1:最も不満足) を与えてもらう。
- (3) 満足度評価に基づいて、被験者一人あたり計200個 (=20コンテキストパターン \times 10飲食店データ) の学習データからなる学習データセットを作成する。

例として、被験者 A が設定したコンテキストパターンを図21に示す。図21(a)は、被験者 A が設定した20個のコンテキストパターンであり、図21(b)は、一つめのコンテキストパターン C_1 をベクトルで表現した例である。なお、連続値型パラメタである[時刻]および[予算]に関しては、本実験では、これらの設定を容易にするため、表7に示すように、あらかじめ取り得る値を用意しておき、この中から被験者に選んでもらうようにした。

被験者 A が作成した学習データセットの一部を抜粋したものを図22(a)に示す。図22(a)のように、各コンテキストパターンにおいて10個ずつの学習データが作成されている。図22(b)のように、一つの学習データは、

(a) 被験者Aが設定した実験用コンテキストパターン

コンテキストパターン	時刻	休日	休日前	予算(円)	同伴者
* コンテキストC ₁	8:00	true	false	1000	なし
コンテキストC ₂	19:00	false	false	2000	家族
コンテキストC ₃	8:00	true	false	1000	家族
* コンテキストC ₄	19:00	true	true	5000	恋人
コンテキストC ₅	12:00	true	false	2000	恋人
コンテキストC ₆	19:00	true	true	5000	友人
コンテキストC ₇	19:00	false	false	1000	なし
* コンテキストC ₈	12:00	false	false	1000	友人
コンテキストC ₉	19:00	true	false	2000	友人
コンテキストC ₁₀	12:00	false	false	1000	なし
コンテキストC ₁₁	12:00	true	false	2000	友人
コンテキストC ₁₂	19:00	false	false	2000	恋人
コンテキストC ₁₃	19:00	true	true	2000	なし
* コンテキストC ₁₄	12:00	true	false	2000	家族
コンテキストC ₁₅	19:00	false	true	3000	恋人
コンテキストC ₁₆	19:00	false	false	2000	友人
* コンテキストC ₁₇	19:00	true	true	3000	家族
コンテキストC ₁₈	19:00	true	false	3000	恋人
コンテキストC ₁₉	12:00	true	false	2000	なし
コンテキストC ₂₀	8:00	false	false	1000	家族

*は評価用データセット作成時に選ばれたコンテキストパターン。

(b) コンテキストのベクトル表現
(コンテキストC₁の場合)

コンテキストパターン	時刻	休日	休日前	予算(円)	同伴者
コンテキストC ₁	8:00	true	false	1000	なし



$$\begin{aligned}
 C_1 &= \{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8\} \\
 &= \{0.333, 1, 0, 0.1, 1, 0, 0, 0\}
 \end{aligned}$$

図 21 実験用コンテキストパターンの設定例 (被験者 A の場合)

- ユーザ満足度 : class ({5, 4, 3, 2, 1})
- 評価対象のアイテムパラメタ : (x₁, x₂, ..., x₁₄)
- 評価時のコンテキストパラメタ : (c₁, c₂, ..., c₈)

により構成されている .

フェーズ 2 : 被験者の評価用データセットの作成

つづいて , 学習データセットとは別に , 以下の手順にしたがって , 各被験者の評価用データセットを作成する .

(a) 被験者Aの学習データセット (一部抜粋) (全200データ)

no	class	x ₁	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈	x ₉	x ₁₀	x ₁₁	x ₁₂	x ₁₃	x ₁₄	c ₁	c ₂	c ₃	c ₄	c ₅	c ₆	c ₇	c ₈	
1	1	1	0	0	0	0	0	0	0	0	0	0	0	0.100	0.1	0.333	1	0	0.1	1	0	0	0	
2	5	0	0	0	0	1	0	0	0	0	0	0	0	0.060	0.5	0.333	1	0	0.1	1	0	0	0	
3	1	0	1	0	0	0	0	0	0	0	0	0	0	0.440	0.1	0.333	1	0	0.1	1	0	0	0	
4	1	0	0	0	0	0	1	0	0	0	0	0	0	0.300	0.3	0.333	1	0	0.1	1	0	0	0	
5	1	0	0	0	0	0	0	1	0	0	0	0	0	0.280	0.3	0.333	1	0	0.1	1	0	0	0	
6	3	0	0	0	0	0	0	0	0	0	0	1	0	0.100	0.1	0.333	1	0	0.1	1	0	0	0	
7	1	1	0	0	0	0	0	0	0	0	0	0	0	0.350	0.3	0.333	1	0	0.1	1	0	0	0	
8	1	0	0	0	0	0	0	0	1	0	0	0	0	0.350	0.3	0.333	1	0	0.1	1	0	0	0	
9	1	0	0	0	1	0	0	0	0	0	0	0	0	0.265	0.1	0.333	1	0	0.1	1	0	0	0	
10	1	0	0	1	0	0	0	0	0	0	0	0	0	0.110	0.5	0.333	1	0	0.1	1	0	0	0	
11	1	0	0	1	0	0	0	0	0	0	0	0	0	0.380	0.4	0.792	0	0	0.2	0	1	0	0	
12	2	0	0	0	0	0	0	0	0	0	0	1	0	0.100	0.1	0.792	0	0	0.2	0	1	0	0	
13	1	0	0	0	0	0	0	0	0	1	0	0	0	0.300	0.5	0.792	0	0	0.2	0	1	0	0	
14	5	1	0	0	0	0	0	0	0	0	0	0	0	0.120	0.1	0.792	0	0	0.2	0	1	0	0	
15	3	1	0	0	0	0	0	0	0	0	0	0	0	0.350	0.3	0.792	0	0	0.2	0	1	0	0	
16	1	0	0	0	1	0	0	0	0	0	0	0	0	0.350	0.5	0.792	0	0	0.2	0	1	0	0	
17	1	0	0	0	0	0	1	0	0	0	0	0	0	0.300	0.2	0.792	0	0	0.2	0	1	0	0	
18	2	0	0	0	0	1	0	0	0	0	0	0	0	0.100	0.1	0.792	0	0	0.2	0	1	0	0	
19	2	0	1	0	0	0	0	0	0	0	0	0	0	0.300	0.1	0.792	0	0	0.2	0	1	0	0	
20	1	0	0	0	0	0	0	0	1	0	0	0	0	0.280	0.2	0.792	0	0	0.2	0	1	0	0	
21	1	0	0	0	0	0	0	0	0	1	0	0	0	0.200	0.3	0.333	1	0	0.1	0	1	0	0	
22	1	0	0	0	0	0	0	0	1	0	0	0	0	0.280	0.2	0.333	1	0	0.1	0	1	0	0	
23	1	0	0	0	1	0	0	0	0	0	0	0	0	0.300	0.3	0.333	1	0	0.1	0	1	0	0	
24	1	0	0	0	0	1	0	0	0	0	0	0	0	0.370	0.3	0.333	1	0	0.1	0	1	0	0	
25	1	0	0	1	0	0	0	0	0	0	0	0	0	0.700	0.1	0.333	1	0	0.1	0	1	0	0	
26	4	0	0	0	0	0	0	0	0	0	0	1	0	0.050	0.1	0.333	1	0	0.1	0	1	0	0	
27	1	1	0	0	0	0	0	0	0	0	0	0	0	0.800	0.5	0.333	1	0	0.1	0	1	0	0	
28	5	0	0	0	0	1	0	0	0	0	0	0	0	0.100	0.1	0.333	1	0	0.1	0	1	0	0	
29	1	0	1	0	0	0	0	0	0	0	0	0	0	0.100	0.1	0.333	1	0	0.1	0	1	0	0	
30	1	1	0	0	0	0	0	0	0	0	0	0	0	0.160	0.1	0.333	1	0	0.1	0	1	0	0	
31	5	0	0	0	0	0	0	0	0	1	0	0	0	0.300	0.3	0.792	1	1	0.5	0	0	1	0	
32	4	0	0	0	0	0	0	0	1	0	0	0	0	0.280	0.2	0.792	1	1	0.5	0	0	1	0	
33	1	0	0	0	0	0	0	0	0	0	0	1	0	0.100	0.1	0.792	1	1	0.5	0	0	1	0	
34	4	0	0	0	0	0	1	0	0	0	0	0	0	0.250	0.3	0.792	1	1	0.5	0	0	1	0	
35	1	0	0	0	0	1	0	0	0	0	0	0	0	0.085	0.1	0.792	1	1	0.5	0	0	1	0	
36	1	0	0	1	0	0	0	0	0	0	0	0	0	0.110	0.5	0.792	1	1	0.5	0	0	1	0	
37	4	0	0	0	1	0	0	0	0	0	0	0	0	0.350	0.2	0.792	1	1	0.5	0	0	1	0	
38	3	0	1	0	0	0	0	0	0	0	0	0	0	0.450	0.5	0.792	1	1	0.5	0	0	1	0	
39	1	1	0	0	0	0	0	0	0	0	0	0	0	0.100	0.2	0.792	1	1	0.5	0	0	1	0	
40	3	1	0	0	0	0	0	0	0	0	0	0	0	0.250	0.3	0.792	1	1	0.5	0	0	1	0	
41	1	0	0	0	0	0	1	0	0	0	0	0	0	0.150	0.3	0.500	1	0	0.2	0	0	1	0	
42	1	0	0	0	0	0	0	0	1	0	0	0	0	0.350	0.5	0.500	1	0	0.2	0	0	1	0	
43	1	0	0	0	0	0	0	0	0	1	0	0	0	0.250	0.5	0.500	1	0	0.2	0	0	1	0	
44	3	1	0	0	0	0	0	0	0	0	0	0	0	0.100	0.1	0.500	1	0	0.2	0	0	1	0	
45	2	0	0	1	0	0	0	0	0	0	0	0	0	0.080	0.5	0.500	1	0	0.2	0	0	1	0	
46	4	0	0	0	0	1	0	0	0	0	0	0	0	0.120	0.1	0.500	1	0	0.2	0	0	1	0	
47	1	0	0	0	0	0	0	0	0	0	0	1	0	0.350	0.5	0.500	1	0	0.2	0	0	1	0	
48	2	0	0	0	1	0	0	0	0	0	0	0	0	0.350	0.5	0.500	1	0	0.2	0	0	1	0	
49	2	0	1	0	0	0	0	0	0	0	0	0	0	0.300	0.4	0.500	1	0	0.2	0	0	1	0	
50	1	1	0	0	0	0	0	0	0	0	0	0	0	0.450	0.5	0.500	1	0	0.2	0	0	1	0	
...
200	1	1	0	0	0	0	0	0	0	0	0	0	0	0.600	0.3	0.333	0	0	0.1	0	1	0	0	

(b) 学習データの説明 (データNo.1の場合)

no	class	x ₁	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈	x ₉	x ₁₀	x ₁₁	x ₁₂	x ₁₃	x ₁₄	c ₁	c ₂	c ₃	c ₄	c ₅	c ₆	c ₇	c ₈
1	1	1	0	0	0	0	0	0	0	0	0	0	0	0.100	0.1	0.333	1	0	0.1	1	0	0	0

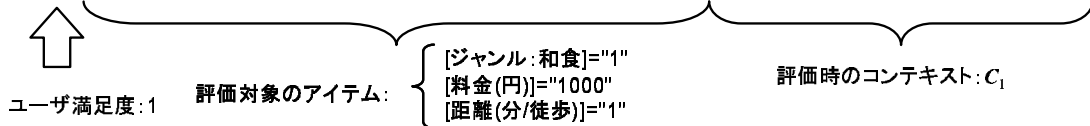


図 22 実験で用いた学習データセット (被験者 A の場合)

- (1) 学習データ作成時に設定した20個のコンテキストパターン $\{C_1, C_2, \dots, C_{20}\}$ の中から，パラメタの値が可能な限りばらつくように5個のコンテキストパターンを選ぶ．
- (2) 被験者には，一つのコンテキストパターンあたり，データセットの中から無作為に抽出された40件の飲食店データに対し，学習データ作成時と同様に，5段階の満足度評価を与えてもらう．
- (3) 満足度評価に基づいて，一つのコンテキストパターンあたり40個のアイテムデータからなる評価用データセットを作成する．最終的に，被験者一人あたり，各コンテキストパターンに対応した5個の評価用データセットが作成される．

被験者 A の例では，図 21 に * を示したように， $\{C_1, C_4, C_8, C_{14}, C_{17}\}$ の5個のコンテキストパターンを選んでいる．被験者 A が作成した評価用データセットを図 23 に示す．図 23 は，上記で示した5個のコンテキストパターンのうち， C_1 に対応した評価用データセットを示している．

フェーズ3：評価用データセットに対するランキング処理

各コンテキストパターンにおける評価用データセットに対し，ランキング処理を行う．

5.4.3 項で述べる比較実験においては，(I) 提案方式の他に，以下の二つのランキング方式を比較対象とする．

(II) 機械学習手法による出力値に基づくランキング方式

(III) 多クラス分類手法に基づくランキング方式

本実験では，(II) の方式として，5.2 節で説明した，SVM の識別平面からの距離に基づくランキング方法 [72][73] を，(III) の方式として，Multi-Class SVM[8][9] に基づくランキング方法をそれぞれ取り上げる．いずれも SVM[18] をベースとしたランキング方法であるが，これは，SVM が現在知られている手法の中で，最も認識性能の高い学習モデルの一つである [82] という根拠に基づく．

被験者Aの評価用データセット (コンテキスト C_1 におけるデータセット)

no	class	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8
1	1	0	0	1	0	0	0	0	0	0	0	0	0	0.300	0.5	0.333	1	0	0.1	1	0	0	0
2	3	1	0	0	0	0	0	0	0	0	0	0	0	0.100	0.1	0.333	1	0	0.1	1	0	0	0
3	1	0	0	0	1	0	0	0	0	0	0	0	0	0.360	0.2	0.333	1	0	0.1	1	0	0	0
4	1	0	0	0	0	0	1	0	0	0	0	0	0	0.250	0.4	0.333	1	0	0.1	1	0	0	0
5	5	0	0	0	0	1	0	0	0	0	0	0	0	0.063	0.2	0.333	1	0	0.1	1	0	0	0
6	3	0	0	0	0	0	0	0	0	0	0	1	0	0.100	0.1	0.333	1	0	0.1	1	0	0	0
7	1	0	0	0	0	0	0	0	0	1	0	0	0	0.300	0.3	0.333	1	0	0.1	1	0	0	0
8	3	1	0	0	0	0	0	0	0	0	0	0	0	1.000	0.2	0.333	1	0	0.1	1	0	0	0
9	1	0	1	0	0	0	0	0	0	0	0	0	0	0.400	0.5	0.333	1	0	0.1	1	0	0	0
10	1	0	0	0	0	0	0	0	1	0	0	0	0	0.300	0.5	0.333	1	0	0.1	1	0	0	0
11	1	0	1	0	0	0	0	0	0	0	0	0	0	0.300	0.5	0.333	1	0	0.1	1	0	0	0
12	1	0	0	0	0	0	0	0	0	0	1	0	0	0.250	0.3	0.333	1	0	0.1	1	0	0	0
13	3	1	0	0	0	0	0	0	0	0	0	0	0	0.100	0.1	0.333	1	0	0.1	1	0	0	0
14	1	0	0	0	0	1	0	0	0	0	0	0	0	0.320	0.2	0.333	1	0	0.1	1	0	0	0
15	1	0	0	0	1	0	0	0	0	0	0	0	0	0.350	0.1	0.333	1	0	0.1	1	0	0	0
16	1	0	0	0	0	0	0	0	1	0	0	0	0	0.250	0.1	0.333	1	0	0.1	1	0	0	0
17	5	0	0	0	0	1	0	0	0	0	0	0	0	0.070	0.1	0.333	1	0	0.1	1	0	0	0
18	4	0	0	0	0	0	0	0	0	0	0	0	1	0.050	0.1	0.333	1	0	0.1	1	0	0	0
19	3	0	0	1	0	0	0	0	0	0	0	0	0	0.085	0.1	0.333	1	0	0.1	1	0	0	0
20	1	1	0	0	0	0	0	0	0	0	0	0	0	0.350	0.6	0.333	1	0	0.1	1	0	0	0
21	3	0	0	0	0	0	0	0	0	0	0	1	0	0.100	0.1	0.333	1	0	0.1	1	0	0	0
22	1	0	0	1	0	0	0	0	0	0	0	0	0	0.300	0.5	0.333	1	0	0.1	1	0	0	0
23	1	0	1	0	0	0	0	0	0	0	0	0	0	0.480	0.1	0.333	1	0	0.1	1	0	0	0
24	5	0	0	0	0	1	0	0	0	0	0	0	0	0.055	0.1	0.333	1	0	0.1	1	0	0	0
25	1	0	0	0	0	0	0	0	0	1	0	0	0	0.300	0.1	0.333	1	0	0.1	1	0	0	0
26	3	1	0	0	0	0	0	0	0	0	0	0	0	0.100	0.1	0.333	1	0	0.1	1	0	0	0
27	1	0	0	0	1	0	0	0	0	0	0	0	0	0.350	0.2	0.333	1	0	0.1	1	0	0	0
28	1	0	0	0	0	0	1	0	0	0	0	0	0	0.380	0.3	0.333	1	0	0.1	1	0	0	0
29	1	1	0	0	0	0	0	0	0	0	0	0	0	0.600	0.5	0.333	1	0	0.1	1	0	0	0
30	1	0	0	0	0	0	0	0	1	0	0	0	0	0.300	0.5	0.333	1	0	0.1	1	0	0	0
31	1	0	0	0	0	0	0	0	0	1	0	0	0	0.300	0.3	0.333	1	0	0.1	1	0	0	0
32	1	0	1	0	0	0	0	0	0	0	0	0	0	0.250	0.3	0.333	1	0	0.1	1	0	0	0
33	1	0	0	0	1	0	0	0	0	0	0	0	0	0.440	0.1	0.333	1	0	0.1	1	0	0	0
34	1	0	0	0	0	0	0	0	0	0	0	1	0	0.350	0.5	0.333	1	0	0.1	1	0	0	0
35	1	0	0	0	0	0	0	0	1	0	0	0	0	0.250	0.1	0.333	1	0	0.1	1	0	0	0
36	3	1	0	0	0	0	0	0	0	0	0	0	0	0.130	0.5	0.333	1	0	0.1	1	0	0	0
37	1	0	0	1	0	0	0	0	0	0	0	0	0	0.280	0.2	0.333	1	0	0.1	1	0	0	0
38	5	0	0	0	0	1	0	0	0	0	0	0	0	0.060	0.5	0.333	1	0	0.1	1	0	0	0
39	1	1	0	0	0	0	0	0	0	0	0	0	0	0.350	0.1	0.333	1	0	0.1	1	0	0	0
40	2	0	0	0	0	0	1	0	0	0	0	0	0	0.150	0.3	0.333	1	0	0.1	1	0	0	0

評価時のコンテキスト: C_1

図 23 実験で用いた評価用データセット (被験者 A の場合)

なお、いずれのランキング方式においても、学習データセットおよび評価用データセットは、それぞれ本項「フェーズ1」および「フェーズ2」で説明したものをを用いる。

本項では、提案方式によるランキング手順と併せ、上記二つの方式によるランキング手順についてそれぞれ説明する。

(I) 提案方式によるランキング

当該コンテキストに関連する学習データの特定

5.3.2 項「当該コンテキストに関連する学習データの特定」で述べた説明にしたがって、以下の手順により、当該コンテキストに関連する学習データを特定する。

- (1) 被験者の学習データセットに基づき、各コンテキストパラメタ c_i について、情報利得を算出する。
- (2) 情報利得の最も大きいコンテキストパラメタの値に基づいて、当該コンテキストに関連する学習データを特定する。

被験者 A の学習データセットを対象にしたとき、当該コンテキストパターンが C_1 の場合の各コンテキストパラメタ c_i の情報利得を求める。

例として、コンテキストパラメタ c_4 、すなわち [予算] パラメタの情報利得を求める。被験者 A の学習データセットより、式 (5), (7) で用いる各値は、次のようになった。

$$\begin{array}{lll}
 N = 200 & n^+ = 77 & n^- = 123 \\
 n|_{c_4 \approx 0.1} = 60 & n^+|_{c_4 \approx 0.1} = 17 & n^-|_{c_4 \approx 0.1} = 43 \\
 n|_{c_4 \neq 0.1} = 140 & n^+|_{c_4 \neq 0.1} = 60 & n^-|_{c_4 \neq 0.1} = 80 \quad (13)
 \end{array}$$

なお、提案方式では、class が 5 段階のうち 3 以上のデータを正データ、2 以下のデータを負データとして扱っている。上記の値を、それぞれ式 (5), (7) に代入す

表 8 各コンテキストパラメタに対する情報利得 (コンテキスト C_1 における被験者 A の場合)

コンテキストパラメタ	情報利得
c_1 時刻	0.0076
c_2 休日	0.0005
c_3 休日前	0.0085
c_4 予算	0.0138
$c_{5\sim 8}$ 同伴者	0.0039

ると,

$$E_1 = 0.9615 \quad (14)$$

$$E_2 = 0.9477 \quad (15)$$

を得る. したがって, 式 (4) より, $Gain_4$ は,

$$Gain_4 = 0.9615 - 0.9477 = 0.0138 \quad (16)$$

となる. 同様に, 各コンテキストパラメタ c_i の情報利得は, 表 8 のとおりとなった.

この中で, コンテキストパラメタ c_4 , すなわち [予算] の情報利得が最も大きかったため, [予算] パラメタを着目するコンテキストパラメタとして選定する.

したがって, この c_4 の値に基づいて, 学習データを特定する. つまり, $0.1 - \alpha \leq c_4 \leq 0.1 + \alpha$ ($\alpha = 0$) となる, 学習データを選べば良いため, 被験者 A の学習データセットから特定した学習データ数は 60 個となる.

なお, 連続値型パラメタの情報利得算出時に用いる α (5.3.2 項参照) は, $\alpha = 0$ とした. これは, 本実験で用いたすべての被験者のデータセットにおいては, $\alpha = 0$ で十分な学習データ数を確保できるためである.

ここでいう, 十分な学習データ数は事前に行った予備実験により決定した. 図 24 は, 予備実験において, 学習データ数と未知データに対するモデルの識別性能の関係を示したものである. データセットとしては, 本実験で用いた被験者の学習データセット, 未知データとしての評価用データセットを用いた. 図 24 より, 学習データ数が 30 のときに, 既に識別性能が収束しているといえるため, 本実験における十分な学習データ数は 30 と判断した. いずれのデータセットにおい

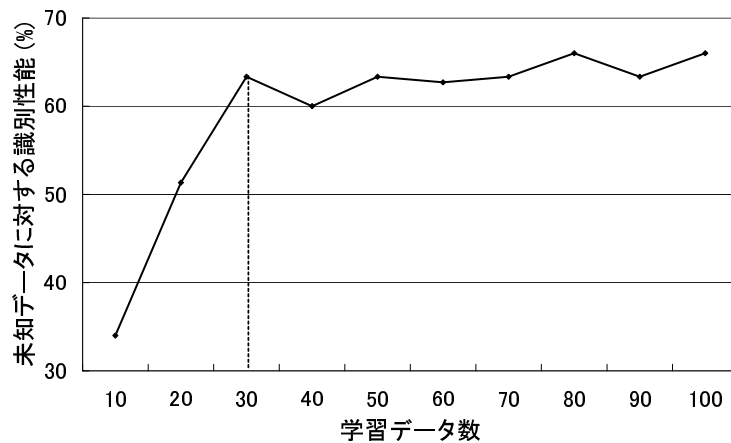


図 24 予備実験：学習データ数と未知データに対する識別性能の関係

ても、 $\alpha = 0$ としたとき、学習データ数 30 は確保できることを確認したため、本実験では $\alpha = 0$ を採用した。

なお、ここでは、着目するコンテキストパラメタとして、最も大きい情報利得を持つコンテキストパラメタを一つ選定する方法を示した。情報利得に基づいて上位いくつかのコンテキストパラメタを選定することにより、より推薦時のコンテキストに限定した学習データを特定することが可能であるが、やはり確保できる学習データ数との兼ね合いが重要となる。後述する、5.4.5 項では、選定するコンテキストパラメタの数について考察を行っている。

被験者の価値判断基準モデルの構築

5.3.2 項「当該コンテキストにおけるユーザの価値判断基準モデルの構築」で述べた説明にしたがって、以下の手順により、当該コンテキストにおける被験者の価値判断基準モデルを構築する。

- (1) 特定された被験者の学習データセットに基づき、各アイテムパラメタ x_i の重要度を算出する。
- (2) 重要度の最も大きいアイテムパラメタ軸における、学習データの分布に基づいて、当該コンテキストにおける被験者の価値判断基準モデルを構築する。

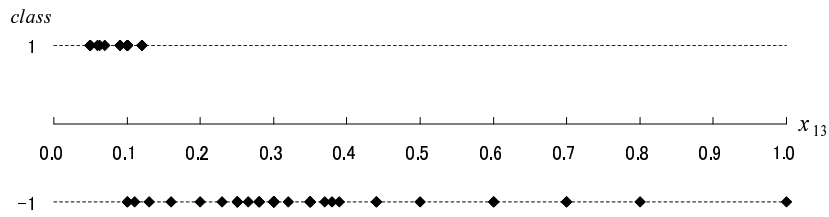


図 25 アイテムパラメタ x_{13} 軸上における正負学習データの分布 (コンテキスト C_1 における被験者 A の場合)

被験者 A の特定された学習データセットに基づいた場合の各アイテムパラメタ x_i の重要度 w_{xi} を求める．例として，アイテムパラメタ x_{13} ，すなわち [料金] パラメタ軸に着目したときの正負データの分布は，図 25 のようになった．図 25 は，横軸にアイテムパラメタ x_{13} の値，縦軸に学習データのクラスを取っている．

ここで，式 (9), (10), (11), (12) で用いる各値は， $r = 0.14$ のとき，次のようになった．

$$n^+ = 17 \qquad n^- = 43 \qquad (17)$$

$$m_i^+ = 94 \qquad m_i^- = 343 \qquad m_i^* = 134 \qquad (18)$$

これを，それぞれ式 (9), (10), (11), (12) に代入すると，

$$w_{x13} = 0.7044 \qquad (19)$$

となる．なお，ここでは， r を $0 \leq r \leq 1$ に変えながら， w_{x13} を求めていった結果， $r = 0.14$ のときに w_{x13} が最大となったため， $r = 0.14$ とした．

同様に，各アイテムパラメタ x_i の重要度 w_{xi} および r は，表 9 のとおりとなった．この中で，アイテムパラメタ x_{13} ，すなわち [料金] パラメタの重要度が最も大きかったため，被験者 A の価値判断基準モデルは，この [料金] パラメタに基づいて構築される．

図 25 の正負データの分布に基づくと，被験者 A の価値判断基準モデルは，図 26 のようになる (5.3.2 項「ユーザの価値判断基準モデルの構築」参照)．図 26 は，横軸にアイテムパラメタ x_{13} の値，縦軸にその値に依存するランキングスコアを取っている．各棒グラフの上端に示している数値は，ランキングスコアを表す．

表 9 各アイテムパラメタの重要度 (コンテキスト C_1 における被験者 A の場合)

アイテムパラメタ		w_{xi}	r
x_1	ジャンル: 和食	0.3471	0
x_2	ジャンル: 洋食	0.4631	0
x_3	ジャンル: 中華	0.4631	0
x_4	ジャンル: アジア・焼肉・エスニック・無国籍	0.5698	0
x_5	ジャンル: カフェ	0.1912	0
x_6	ジャンル: 和風居酒屋	0.5698	0
x_7	ジャンル: 洋風居酒屋	0.5233	0
x_8	ジャンル: アジア・無国籍居酒屋	0.5465	0
x_9	ジャンル: ダイニングバー・ビアレストラン	0.5233	0
x_{10}	ジャンル: バー・パブ	0.5465	0
x_{11}	ジャンル: スペース・エンターテイメント	0.2794	0
x_{12}	ジャンル: ファミレス・ファーストフード	0.2794	0
x_{13}	料金	0.7044	0.14
x_{14}	距離	0.4690	0

なお，ここでは，着目するアイテムパラメタとして，最も大きい重要度を持つアイテムパラメタを一つ選定する方法を示した．二つ以上のアイテムパラメタを選定した場合には，各アイテムパラメタ軸上における学習データの分布を調べ，それぞれのアイテムパラメタに対する価値判断基準モデルを構築する．そして，ランキングスコアは各モデルに基づいた算出された総和により求められる．後述する，5.4.4 項では，選定するアイテムパラメタの数について考察を行っている．

価値判断基準モデルに基づくランキング

最終的に，各評価用コンテキストパターンにおける価値判断基準モデルに基づき，対応するコンテキストパターンごとの評価用データセット内の各データに対しスコアを算出する．このスコアに基づいてランキングを行う．つまり，評価用データセット内の 40 個のデータに対し，提案方式により 1～40 位までのランク付けを行う．

被験者 A のコンテキスト C_1 における評価用データセットを C_1 における価値判断基準モデルに基づきランキングした結果は，図 27(a) のようになる．なお，図 27(b) に示すように，例えば，データ No.5 の場合，重要アイテムパラメタである x_{13} の値が $0 \leq x_{13} \leq 0.1$ の範囲にあるので，図 26 の $0 \leq x_{13} \leq 0.1$ に対応す

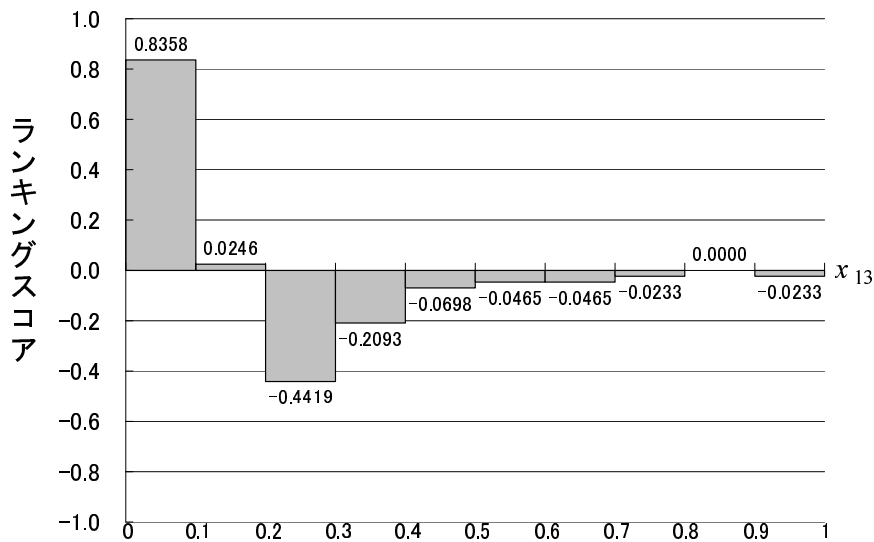


図 26 アイテムパラメタ x_{13} に対するユーザの価値判断基準モデル (コンテキスト C_1 における被験者 A の場合)

るランキングスコアより, 0.8358 が加算される. このランキングスコアを各データに対し算出し, 高いものから上位にランク付けする.

これをコンテキストパターンの異なる 5 個の評価用データに対し行う. この結果を提案方式によるランキング結果とする.

(II) 機械学習手法による出力値に基づくランキング

機械学習手法では, 入力データに基づいて何らかの値を出力し, その出力値に基づいて入力データの識別等を行う. そこで, 単純には, この出力値の大きさや, 目標値との近さなどに基づいて入力データに対するランキング処理を行うことができると考えられる.

SVM[18] では, 次式で表される識別関数に基づいて, 入力データ x を正負クラスに分類する.

$$f(x) = \text{sgn}(g(x)) \quad (20)$$

$$g(x) = \mathbf{w}^t \mathbf{x} + b \quad (21)$$

(a) 提案方式による評価用データセットに対するランキング (コンテキスト C_1 における被験者Aの場合)

no	class	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	score
5	5	0	0	0	0	1	0	0	0	0	0	0	0	0.063	0.2	0.8358
17	5	0	0	0	0	1	0	0	0	0	0	0	0	0.070	0.1	0.8358
24	5	0	0	0	0	1	0	0	0	0	0	0	0	0.055	0.1	0.8358
38	5	0	0	0	0	1	0	0	0	0	0	0	0	0.060	0.5	0.8358
18	4	0	0	0	0	0	0	0	0	0	0	0	1	0.050	0.1	0.8358
19	3	0	0	1	0	0	0	0	0	0	0	0	0	0.085	0.1	0.8358
2	3	1	0	0	0	0	0	0	0	0	0	0	0	0.100	0.1	0.0246
6	3	0	0	0	0	0	0	0	0	0	0	1	0	0.100	0.1	0.0246
13	3	1	0	0	0	0	0	0	0	0	0	0	0	0.100	0.1	0.0246
21	3	0	0	0	0	0	0	0	0	0	0	1	0	0.100	0.1	0.0246
26	3	1	0	0	0	0	0	0	0	0	0	0	0	0.100	0.1	0.0246
36	3	1	0	0	0	0	0	0	0	0	0	0	0	0.130	0.5	0.0246
40	2	0	0	0	0	0	1	0	0	0	0	0	0	0.150	0.3	0.0246
8	3	1	0	0	0	0	0	0	0	0	0	0	0	1.000	0.2	0.0000
29	1	1	0	0	0	0	0	0	0	0	0	0	0	0.600	0.5	-0.0465
9	1	0	1	0	0	0	0	0	0	0	0	0	0	0.400	0.5	-0.0698
23	1	0	1	0	0	0	0	0	0	0	0	0	0	0.480	0.1	-0.0698
33	1	0	0	0	1	0	0	0	0	0	0	0	0	0.440	0.1	-0.0698
1	1	0	0	1	0	0	0	0	0	0	0	0	0	0.300	0.5	-0.2093
3	1	0	0	0	1	0	0	0	0	0	0	0	0	0.360	0.2	-0.2093
7	1	0	0	0	0	0	0	0	0	1	0	0	0	0.300	0.3	-0.2093
10	1	0	0	0	0	0	0	0	1	0	0	0	0	0.300	0.5	-0.2093
11	1	0	1	0	0	0	0	0	0	0	0	0	0	0.300	0.5	-0.2093
14	1	0	0	0	0	0	1	0	0	0	0	0	0	0.320	0.2	-0.2093
15	1	0	0	0	1	0	0	0	0	0	0	0	0	0.350	0.1	-0.2093
20	1	1	0	0	0	0	0	0	0	0	0	0	0	0.350	0.6	-0.2093
22	1	0	0	1	0	0	0	0	0	0	0	0	0	0.300	0.5	-0.2093
25	1	0	0	0	0	0	0	0	0	1	0	0	0	0.300	0.1	-0.2093
27	1	0	0	0	1	0	0	0	0	0	0	0	0	0.350	0.2	-0.2093
28	1	0	0	0	0	0	1	0	0	0	0	0	0	0.380	0.3	-0.2093
30	1	0	0	0	0	0	0	0	1	0	0	0	0	0.300	0.5	-0.2093
31	1	0	0	0	0	0	0	0	0	1	0	0	0	0.300	0.3	-0.2093
34	1	0	0	0	0	0	0	0	0	0	0	1	0	0.350	0.5	-0.2093
39	1	1	0	0	0	0	0	0	0	0	0	0	0	0.350	0.1	-0.2093
4	1	0	0	0	0	0	1	0	0	0	0	0	0	0.250	0.4	-0.4419
12	1	0	0	0	0	0	0	0	0	0	1	0	0	0.250	0.3	-0.4419
16	1	0	0	0	0	0	0	0	1	0	0	0	0	0.250	0.1	-0.4419
32	1	0	1	0	0	0	0	0	0	0	0	0	0	0.250	0.3	-0.4419
35	1	0	0	0	0	0	0	0	1	0	0	0	0	0.250	0.1	-0.4419
37	1	0	0	1	0	0	0	0	0	0	0	0	0	0.280	0.2	-0.4419

(b) スコア算出の例 (データNo.5の場合)

no	class	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	score
5	5	0	0	0	0	1	0	0	0	0	0	0	0	0.063	0.2	0.8358



重要アイテムパラメタである x_{13} の値が0~0.1の間であるから、
 価値判断基準モデルより対応するスコア0.8358を加算。

図 27 提案方式による評価用データセットに対するランキング (被験者 A の場合)

ここで、 w および b は、学習データセットに基づいて決定される。このとき、 $f(x) = +1$ となる x を正クラス、 $f(x) = -1$ となる x を負クラス、としてそれぞれ判別する。

$g(x)$ に着目すると、特徴空間上、 $g(x) = 0$ となる超平面が識別超平面を表す。この識別超平面を境界に、 $g(x) > 0$ となる x を正クラス、 $g(x) < 0$ となる x を負クラス、としてそれぞれ判別される。このとき、単純には、 $g(x) \gg 0$ のとき、特徴空間上で、 x は識別超平面から、より遠くに位置するため、確信的に正クラスに属すと考えることができる。この考えに基づくと、 $g(x)$ の値に基づいたランキング [73] が可能となる。

本実験では、学習データセットとして、提案方式で用いたものと同等のもの(本節「フェーズ1」参照)を用いて学習を行う。ここで、提案方式と同様に、class が5段階のうち3以上のデータを正データ、2以下のデータを負データとして扱う。そして、同じく提案方式で用いたものと同等の、5個のコンテキストパターンにおける評価用データセット(本節「フェーズ2」参照)に対して、 $g(x)$ に基づくランキングを行う。

なお、SVMの実装には、LIBSVM[63][64]を用い、SVMの学習に必要なパラメタ C および γ は、grid-search[63]により学習データセットにとって最適な値を設定した。また、カーネルトリックとして、RBFカーネルを使用した。

(III) 多クラス分類手法に基づくランキング方式

入力データを多クラスに分類する多クラス分類手法がある。このクラスを、例えば $\{5, 4, 3, 2, 1\}$ のようなランクとして見立てることにより、多クラス分類手法に基づくランキングを行うことができる。

SVMは2クラス分類手法であるが、これを多クラス分類へと拡張したものが Multi-Class SVM[8][9]である。

Multi-Class SVMでは、 k クラス分類問題を k 段階の2クラス分類問題として扱っている。例えば、第 n 段階の分類問題は、クラス n とそれ以外の $k - 1$ 個のクラスとの2クラス分類問題とみなすことができる。

本実験では、被験者満足度として $\{5, 4, 3, 2, 1\}$ の5クラスを用いている。このとき、学習データセット内の各データのクラスを次のように定義する。

- 正クラス {5} vs. 負クラス {4, 3, 2, 1}
- 正クラス {4} vs. 負クラス {5, 3, 2, 1}
- 正クラス {3} vs. 負クラス {5, 4, 2, 1}
- 正クラス {2} vs. 負クラス {5, 4, 3, 1}
- 正クラス {1} vs. 負クラス {5, 4, 3, 2}

上記の2クラスで定義された各学習データセットに基づいて学習モデルを構築する。そして、このモデルに基づき、各評価用データセット内の各データを {5, 4, 3, 2, 1} の5クラスに分類する。この分類結果に基づき、{5, 4, 3, 2, 1} の順にランキングを行う。

なお、Multi-Class SVMにおいても、学習データセットおよび評価用データセットは、提案方式で用いたものと同等のものを用いる。また、SVMの実装は、(II)で示した設定と同様である。

フェーズ4：ランキング結果に対する評価

以上の各手法によるランキング結果を比較することで、提案手法の有効性を評価するのであるが、評価尺度として、DCG (Discounted Cumulative Gain) [83] を用いる。DCGは、多段階の適合性に適した評価尺度であり、次式で定義される。

$$DCG(i) = \begin{cases} G(1) & \text{if } i = 1 \\ DCG(i-1) + \frac{G(i)}{\log_2 i} & \text{otherwise} \end{cases} \quad (22)$$

ここで、 $G(i)$ は、第 i 位にランク付けされたアイテムの適合性を示す。本実験では、被験者による満足度評価として、“5” と与えられたアイテムを“最適”，“4” と与えられたアイテムを“高適合”，“3” と与えられたアイテムを“適合”，そ

れ以外を“不適合”とし， $G(i)$ を次のように定義する．

$$G(i) = \begin{cases} 5 & \text{最適合の場合} \\ 4 & \text{高適合の場合} \\ 3 & \text{適合の場合} \\ 0 & \text{不適合} \end{cases} \quad (23)$$

本実験では，ランキング結果の上位 5, 10, 15, 20 件に対し，DCG を算出する．この DCG の値が大きいほど有効なランキングであることを示す．

ランキング手法などの評価尺度には，アイテムの適合・不適合だけに基づいて算出される適合率が用いられることが多いが，以上のように多段階の適合性を考慮して定義される DCG を用いることで，被験者の満足の程度をより反映させた評価が行うことができる．

5.4.3 各ランキング方式との比較

図 28 は，各ランキング方式によるランキング結果から上位 5, 10, 15, 20 件に対する DCG を算出した結果を示している．いずれの件数においても，提案方式の方が既存のランキング方式に対し，DCG が上回っていることが確認できた．

以下，まず比較対象の各ランキング方式が，提案方式よりも DCG が下回った根拠について考察する．特に，図 29 を踏まえて，各方式により判定されたランキングがどれだけユーザの嗜好を反映できているか，という観点から考察を行う．図 29 は，被験者 A のコンテキストパターン C_1 における評価用データセットに対し，各方式によるランキング結果を示している．それぞれ，データ No.，被験者 A が与えた満足度，各方式によるスコアを示している．

まず，図 29(II) より，SVM の識別平面からの距離に基づく方法では， $g(x)$ の符号により，データの正負クラスの判別はある程度行えているものの， $g(x)$ の値が大きいほど，被験者満足度が高いという結果にはならなかった．この結果より，ユーザの嗜好に関しては， $g(x)$ の符号により，満足か不満足かの判別は行えるが， $g(x)$ の値そのものがユーザの嗜好の程度を表しているとはいえない．したがって， $g(x)$ の値に基づいてランキングを行うことは適切ではないと考えられる．

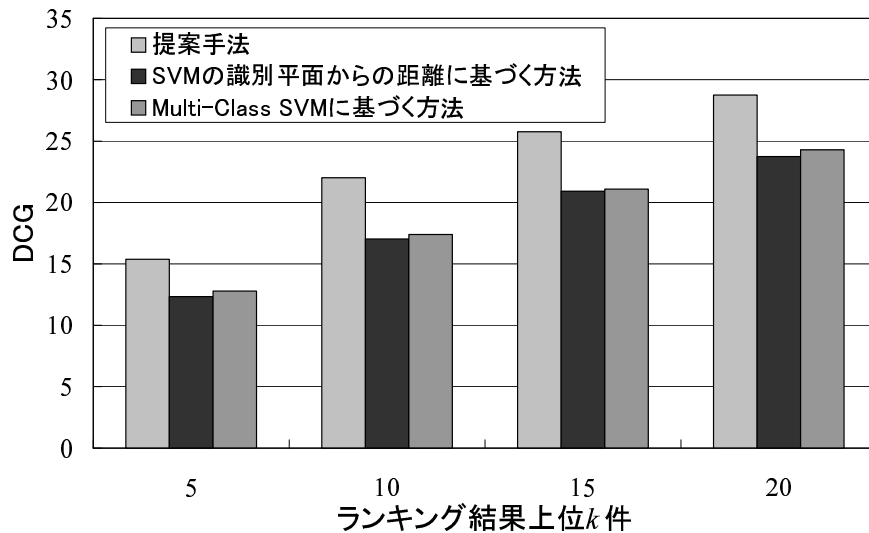


図 28 ランキング手法の比較

つづいて、図 29(III) より、Multi-Class SVM に基づく方法では、最上位クラス（この例ではクラス “4”）と判定されたデータは 6 個あったが、その中には被験者満足度 “3” や “1” と評価されたデータも混在している。このように、Multi-Class SVM により判定された結果が、被験者が与えた満足度と一致していない。これは、学習データセット作成時に評価付けされた各データに対する満足度が、一貫した評価基準に基づいて付けられたものではないということが原因であると考えられる。例えば、満足度 “5” と評価付けされたデータは、他の “5” と評価付けされたどのデータとも、一貫して同等のものであることが保証されていなければならない。しかし、実際にはユーザによる満足度評価は曖昧なものであり、あるコンテキストにおいて “5” と評価付けされたデータと、また別のコンテキストにおいて “5” と評価付けされたデータは、それぞれ違ったコンテキスト下における相対的な価値観に基づくものになる。このため、これらのデータが必ずしもユーザにとって同等のランクに属するとは限らない。このようにコンテキストが異なる場合に、ユーザの価値判断基準も変わってしまう場合、多段階に分割されたユーザ満足度を、そのままランキングのためのクラスとして扱うことは適切でないと考えられる。このことは、Multi-Class SVM に限ったことではなく、一般に多ク

(I) 提案方式による
ランキング

no	被験者 満足度	提案方式 によるスコア
38	5	0.8358
24	5	0.8358
17	5	0.8358
5	5	0.8358
18	4	0.8358
19	3	0.8358
36	3	0.0246
26	3	0.0246
21	3	0.0246
13	3	0.0246
6	3	0.0246
2	3	0.0246
40	2	0.0246
8	3	0.0000
29	1	-0.0465
9	1	-0.0698
23	1	-0.0698
33	1	-0.0698
20	1	-0.2093
10	1	-0.2093
11	1	-0.2093
7	1	-0.2093
15	1	-0.2093
3	1	-0.2093
14	1	-0.2093
25	1	-0.2093
1	1	-0.2093
27	1	-0.2093
22	1	-0.2093
28	1	-0.2093
31	1	-0.2093
30	1	-0.2093
39	1	-0.2093
34	1	-0.2093
37	1	-0.4419
32	1	-0.4419
4	1	-0.4419
12	1	-0.4419
16	1	-0.4419
35	1	-0.4419

(II) SVMの識別平面
に基づくランキング

no	被験者 満足度	$g(x)$
38	5	3.050
36	3	2.792
37	1	2.333
32	1	2.259
26	3	1.702
24	5	1.687
20	1	1.215
18	4	1.138
21	3	0.984
19	3	0.965
17	5	0.949
13	3	0.870
10	1	0.650
8	3	0.250
6	3	0.130
4	1	0.065
2	3	0.000
9	1	-0.580
11	1	-0.790
7	1	-0.850
12	1	-0.875
15	1	-0.915
5	5	-0.929
3	1	-0.992
14	1	-1.015
23	1	-1.042
16	1	-1.115
25	1	-1.196
1	1	-1.200
27	1	-1.321
22	1	-1.620
29	1	-1.631
28	1	-1.770
31	1	-1.961
30	1	-2.034
33	1	-2.123
35	1	-2.243
39	1	-2.451
34	1	-2.497
40	2	-3.209

(III) Multi-Class SVM
に基づくランキング

no	被験者 満足度	Multi-Class SVMにより 判定されたクラス
17	5	4
24	5	4
13	3	4
19	3	4
32	1	4
37	1	4
5	5	3
6	3	3
8	3	3
21	3	3
26	3	3
4	1	3
10	1	3
20	1	3
18	4	2
38	5	1
2	3	1
36	3	1
40	2	1
1	1	1
3	1	1
7	1	1
9	1	1
11	1	1
12	1	1
14	1	1
15	1	1
16	1	1
22	1	1
23	1	1
25	1	1
27	1	1
28	1	1
29	1	1
30	1	1
31	1	1
33	1	1
34	1	1
35	1	1
39	1	1

図 29 各方式による評価用データセットに対するランキング (コンテキスト C_1 における被験者 A の場合)

ラス分類手法を用いた場合にもいえることであると考える。

以上に対し，図 29(I) より，提案方式では，算出されたランキングスコアと被験者満足度とがほぼ一致していることが分かる．提案方式では，まず推薦時のコンテキストにより限定された学習データを特定したうえで，さらにユーザが重要視するアイテムパラメタ軸に着目したときに，その軸上での満足データおよび不満足データの分布に基づいて，厳密にランキングスコアを算出している．このため，推薦時のコンテキストにおけるユーザの価値判断基準をより忠実に反映したランキングを提供することが可能である．結果として，本実験において提案方式が優位性を示したといえる．

5.4.4 ユーザの価値判断基準モデル構築の対象とするアイテムパラメタ選定に関する考察

5.3.2 項「当該コンテキストにおけるユーザの価値判断基準モデルの構築」で述べたように，ユーザの価値判断基準モデル構築の対象とするアイテムパラメタは，ユーザのアイテムパラメタに対する重要度に基づいて選定される．本実験では，重要度の高い順にアイテムパラメタを選定し，このパラメタを対象に価値判断基準モデルを構築したとき，アイテムパラメタの選び方により，ランキング結果に対する DCG の値がどのように変化するか検証する．

図 30 は，アイテムパラメタ選定によるランキング結果に対する DCG の値の違いを示したものである．ここでは，アイテムパラメタに対する重要度に基づき，以下のそれぞれのアイテムパラメタに着目し，被験者の価値判断基準モデルを構築したときのランキング結果に対する DCG を算出している．

- (1) 第 1 位のアイテムパラメタ
- (2) 第 2 位のアイテムパラメタ
- (3) 第 3 位のアイテムパラメタ

この結果から，いずれの件数においても，第 1 位のアイテムパラメタに着目したときが，最も高い DCG を示していることが確認できる．つづいて，第 2 位の

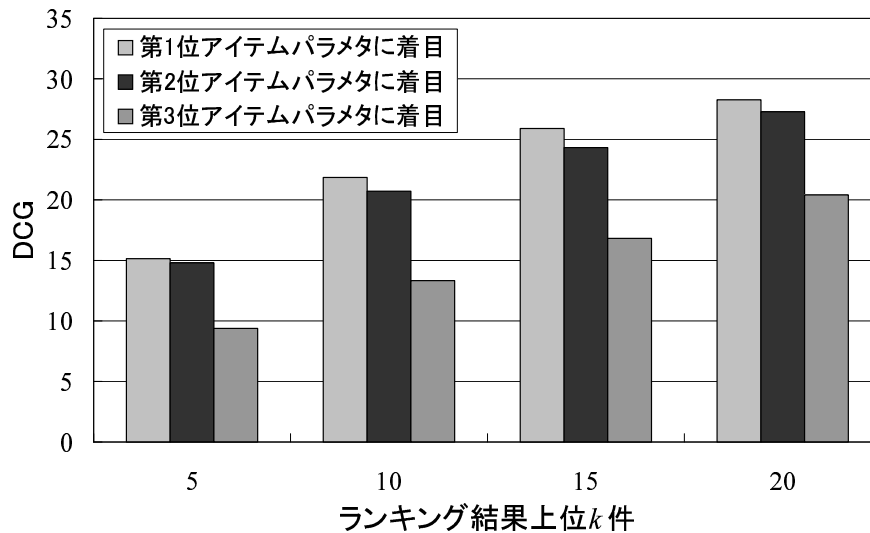


図 30 アイテムパラメタ選定によるランキング結果に対する DCG の値の違い

アイテムパラメタに着目したときも高い DCG を示している。この結果は、重要度が高いアイテムパラメタほど、被験者の価値判断基準モデルが適切に構築されていることを示す。したがって、この結果から本章で定義した重要度の算出方法は妥当であったといえる。

一方で、第 3 位のアイテムパラメタに着目したときには、第 1,2 位のとく比べ、DCG の値が極端に小さくなっている。これは、今回の実験では被験者が飲食店を評価するにあたり、第 3 位のアイテムパラメタに関しては、ほとんど注目していなかったことが原因であると考えられる。実験結果を詳細にみると、第 3 位のアイテムパラメタは、ほとんどの被験者において [距離] パラメタであった。実際に、被験者に対するヒアリング調査結果からも、「[距離] パラメタに関しては、ほとんど注目していなかった」という回答が多かった。この結果からも、被験者が注目していないようなアイテムパラメタが下位に判定されたことから、重要度の算出方法の妥当性が確認できる。

以上のように、ユーザがほとんど注目していないようなアイテムパラメタを対象に価値判断基準モデルを構築することはかえってノイズになることが考えられるため、モデル構築の対象からは除外することが望ましいと考えられる。このよ

うな観点から，モデル構築を行った際，DCG スコアを低下させるようなアイテムパラメタを除外することで，ランキング結果に対するユーザ満足度をより向上させることができると考えられる．

5.4.5 ユーザの価値判断基準に影響を及ぼすコンテキストパラメタ選定に関する考察

5.3.2項「当該コンテキストに関連する学習データの特定」で述べたように，ユーザコンテキストに関連する学習データの特定にあたり，着目するコンテキストパラメタは，コンテキストパラメタのユーザの価値判断基準への影響度に基づいて選定される．本実験では，この影響度の高い順にコンテキストパラメタを選定し，このパラメタに基づいて，学習データの特定を行ったとき，コンテキストパラメタの選び方により，ランキング結果に対する DCG の値がどのように変化するかを検証する．

図 31 は，コンテキストパラメタ選定によるランキング結果に対する DCG の値の違いを示したものである．ここでは，コンテキストパラメタのユーザの価値判断基準への影響度に基づき，以下のそれぞれのコンテキストパラメタに着目し，被験者の価値判断基準モデルを構築したときのランキング結果に対する DCG を算出している．

- (1) コンテキストパラメタに着目しない
- (2) 第 1 位のコンテキストパラメタ
- (3) 第 1,2 位のコンテキストパラメタ
- (4) 第 1,2,3 位のコンテキストパラメタ

ここで，コンテキストパラメタに着目しない場合は，被験者の全学習データに基づいて価値判断基準モデルが構築される．この場合，コンテキストパラメタに着目した場合に比べ，ランキング結果に対する DCG の値が小さくなっている．この結果は，コンテキスト依存のランキングを行うためには，ユーザコンテキスト

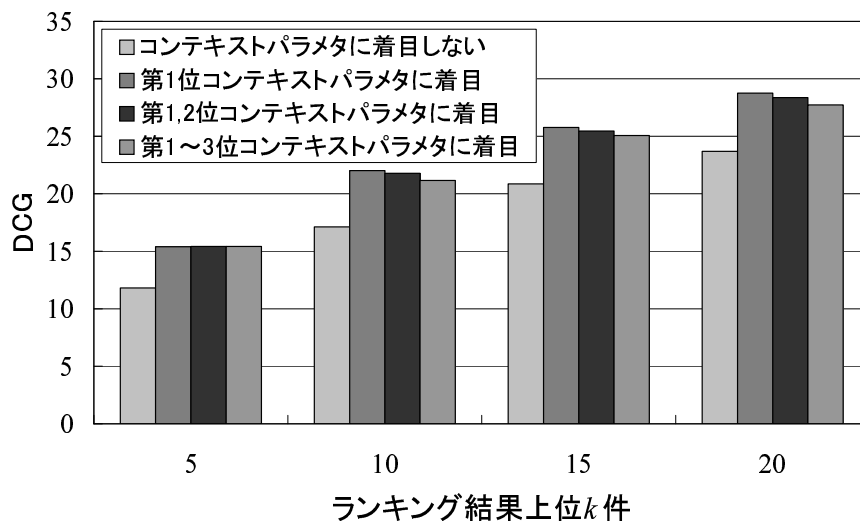


図 31 コンテキストパラメタ選定によるランキング結果に対する DCG の値の違い

に関連する学習データを特定したうえで、価値判断基準モデルを構築することが重要であるということを裏付けている。

全体的にみると、第1位のコンテキストパラメタに着目したときに比べ、第1,2位、第1~3位と、DCGが減少傾向にある。これは、着目するコンテキストパラメタの数が増えることに伴い、そのコンテキストパラメタに関連する学習データ数が減少してしまうため、適切な価値判断基準モデルが作成できていないことが原因であると考えられる。

したがって、各コンテキストパラメタの影響度を考慮し、DCGスコアが低下しない範囲において、複数のコンテキストパラメタ群を考慮したランキングを行うことが可能となる。これにより、ユーザにとって満足度の高い多様なアイテムの提示が可能になるといえる。

5.5 結言

本章では、ユーザの価値判断基準モデルを定義したうえで、このモデルに基づいたコンテキスト依存型ランキング方式を提案した。

本方式の流れは次のとおりである。

- (1) 学習データからのユーザコンテキストに依存する価値判断基準モデルの構築。
- (2) 価値判断基準モデルに基づく推薦候補アイテム集合に対するランキング。

本章では、また提案方式の有効性を評価するために、飲食店情報推薦を対象とした被験者実験を行った。本実験により得られた成果と知見を以下に示す。

- 提案方式では、ユーザコンテキストに応じて変化するユーザの価値判断基準を適切にモデル化することにより、ユーザの直面しているコンテキストにおいて相応しいランキング結果を提供することができる。
- ユーザのアイテムパラメタに対する重要度に基づいて、ユーザが重要視しているアイテムパラメタを適切に選定することにより、ランキング結果に対するユーザ満足度をより向上させることができる。
- コンテキストパラメタのユーザ嗜好に対する影響度に基づいて、影響の大きいコンテキストパラメタのみに着目し学習データを特定することにより、ユーザの直面しているコンテキストにおいて、より相応しいランキング結果を提供することができる。

6. ユーザコンテキストの時系列性を考慮した情報推薦方式

6.1 緒言

ユーザの嗜好は，ユーザがおかれているその時点での状況だけでなく，ユーザが過去に取った行動や将来予定している行動にも影響を受けると考えられる．例えば，「バーで飲んだ後はカラオケで歌いたい」，「1時間後に友達と遊ぶ約束をしているので，近くのカフェで時間をつぶしたい」などといった場面が挙げられる．このため，第3章で述べたユーザコンテキストの定義，すなわち「情報選択過程におけるユーザの嗜好に影響を及ぼす，ユーザのおかれている状況」に則ると，ユーザの嗜好に影響を及ぼすような，過去/未来といったユーザの時系列的な行動履歴もユーザコンテキストの一部であると考えられる．したがって，コンテキスト依存型情報推薦において，過去/未来のユーザの時系列的な行動履歴を考慮に入れることは重要な課題である．

ユーザの時系列的な行動履歴を考慮した情報推薦として，単純には，ユーザの行動パターンに沿った情報推薦 [10][11] が考えられる．例えば，ユーザ A は，日常的に「買物をして，食事を取った後には，カラオケに行く」，すなわち < “買物する” \Rightarrow “食事する” \Rightarrow “カラオケに行く” > という一連の行動パターンを持つとする．この場合，ある日，ユーザ A が「買物をして，食事を取った」とすると，情報推薦システムは，ユーザ A の行動パターンを考慮して，ユーザ A に近くのカラオケ店を推薦することにより，ユーザ A を支援することができる．

この例では，ユーザ A は，

- “買物する” という行動に影響を受けて “カラオケに行く” ．
- “食事する” という行動に影響を受けて “カラオケに行く” ．
- “買物する” かつ “食事する” という行動に影響をうけて “カラオケに行く”

のいずれかであることが分かる．このように，ユーザが過去のどの行動に影響を受けて，次の行動を決定しているかを調べることによって，ユーザの次の行動が

予測でき、ユーザへの情報推薦が可能になると考えられる。基本的には、ユーザの行動履歴データベースから、ユーザの行動パターンを抽出することにより、上記のことを調べることができる。

しかしながら、ユーザの行動パターンも、やはり、ユーザがその時点におかれているコンテキスト（推薦時コンテキストとよぶ）によって変わってくる。例えば、同じ“食事を取った”後であったとしても、

- 「友人といるときには、次にはカラオケに行きたい。」
- 「恋人といるときには、次にはバーで飲みたい。」

といった場面が考えられる。この場合、“友人といる”、“恋人といる”という推薦時コンテキストと、“食事を取った”というユーザの行動履歴を考慮して、推薦情報を決定する必要がある。

そこで、本章で提案する推薦方式では、ユーザの行動履歴データベースから、そのユーザの一般的な行動パターンを抽出するのではなく、推薦時コンテキストに応じた限定的な行動パターンを抽出することで、推薦時コンテキストにおいて、ユーザの過去の行動履歴が次の行動に及ぼす影響の大きさを調べる。得られた影響の大きさから、ユーザの次の行動を予測し、その予測に基づいた推薦を行う。

以下、本章の構成を示す。6.2節では、基礎的事項として、行動パターン抽出手法を説明し、関連研究としてユーザの行動パターンに沿った情報推薦技術を取り上げる。6.3節では、提案方式の基本アイデアとなる部分について説明する。6.4節では、提案方式の実装に、既存の行動パターン抽出手法を適用したときの例を示す。6.5節において、被験者実験を通じて提案方式の有効性を評価し、6.6節で本章をまとめる。

6.2 基礎的事項

6.2.1 行動パターン抽出手法

ユーザの行動パターンとは、ユーザがよく取る一連の行動の流れである。基本的には、ユーザの行動パターンは、ユーザがこれまでに取った行動に関するデータを

時系列に蓄積した行動履歴データベースから抽出される。ユーザの行動履歴データベースから、行動パターンを抽出する手法として、シーケンシャルパターンマイニング手法 [84] を用いた方法や最長共通部分列 (Longest Common Subsequence; LCS) に基づく方法 [11] などが挙げられる。6.4 節では、本論文の提案方式に各手法を適用した例を示す。以下、各手法について説明する。

シーケンシャルパターンマイニング手法を用いた方法

シーケンシャルパターンマイニング手法 [84] は、データマイニング手法 [85] の一つであり、アイテムの発生順序を考慮して頻出パターンを抽出する手法である。顧客の購買行動や自然災害、Web ページのクリックの流れ、株取引などアイテムの発生順序が重要となるケースにおいて有用な情報を抽出することができる [86]。シーケンシャルパターンマイニングを実現するアルゴリズムとして、PrefixSpan[87] や SPADE[88]、PAID[89] などが提案されている。

Agrawal ら [84] は、シーケンシャルパターンマイニングを以下のように定式化している。

非空なアイテムの集合をアイテム集合とよび、アイテム集合 I を (i_1, i_2, \dots, i_m) と表記する。ここで、 i_j は任意のアイテムである。また、順序付けされたアイテム集合のリストをシーケンスとよび、シーケンス S を $\langle I_1, I_2, \dots, I_n \rangle$ と表記する。ここで、 I_j は任意のアイテム集合である。

いま、 $a_1 \subseteq b_{i_1}, a_2 \subseteq b_{i_2}, \dots, a_n \subseteq b_{i_n}$ となるとき、シーケンス $\alpha = \langle a_1, a_2, \dots, a_n \rangle$ は、シーケンス $\beta = \langle b_1, b_2, \dots, b_m \rangle$ に「含まれる」といい、 $\alpha \prec \beta$ と表記する。

シーケンスデータベース Ω は、シーケンスの集合であり、

$$\Omega = \{S_1, S_2, \dots, S_n\} \quad (24)$$

と表記される。ここで、 S_j は任意のシーケンスである。

いま、シーケンス s が S_j に含まれるとき、 S_j は s を「サポートする」という。シーケンスデータベース内の全シーケンスに対し、 s をサポートしているシーケンスの割合を s のサポート値とよび、 $\text{sup}(s)$ と表記する。

シーケンシャルパターンマイニング手法は、シーケンスデータベースの中から、最小サポート値 minsup を満たすシーケンスを抽出するものである。最小サポー

ト値を満たすシーケンスのことを頻出シーケンスとよぶ。

ここで、アイテムをユーザの取った行動として、また頻出シーケンスを行動パターンとして扱うことにより、ユーザの行動履歴データベースから、ユーザの行動パターンを抽出することができる。

最長共通部分列 (Longest Common Subsequence; LCS) に基づく方法

複数のアイテム列の間で共通に含まれる部分列を共通部分列といい、共通部分列の中で最長のものを最長共通部分列 (Longest Common Subsequence) とよび、LCS と略される [90] [91]。いま、アイテム列 $X = abcdbb$ とアイテム列 $Y = cbacbaaba$ があるとすると、このとき、 X と Y の LCS は、 $bcbb$ および $acbb$ となり、LCS の長さは 4 である。

宇根田ら [92] や戸田ら [93] は、Web ページのアクセスログから LCS を抽出する方法を提案している。この手法では、まず Web アクセスログからユーザセッションを切り出す。ユーザセッションとは、WWW におけるアクセス数の単位の一つであり、Web サイトを訪れたユーザがサイト内で行う一連の行動をまとめて 1 セッションという。Web アクセスログから得られた各セッションをシーケンスとして扱い、各シーケンスについて総当りに LCS を抽出する。そして、各 LCS の出現頻度から、高頻度で出現する LCS を抽出する。

山元ら [11] は、LCS を用いた Web ページ推薦手法を提案している。アクセスログから抽出された LCS と現在までのユーザのセッション (アクティブセッション) との照合を行い、ユーザへの推薦候補ページを決定する。ここで、抽出された LCS のうち、全セッションにおける出現回数が $min.Count$ 以上であり、かつ長さ $min.Length$ 以上である LCS の集合を Large LCS 集合とよび、 $LL = \{lcs_1, lcs_2, \dots, lcs_k\}$ で表す。このとき、長さ n のアクティブセッション act_n と LL とを比較し、ユーザへの推薦ページを決定する。

6.2.2 ユーザの行動パターンに沿った情報推薦技術

ユーザの行動パターンに沿った情報推薦技術に関しては、これまでもいくつか行われている。

6.2.1項「最長共通部分列 (Longest Common Subsequence; LCS) に基づく方法」で説明したように、山元ら [11] は、LCS を用いた Web ページ推薦手法を提案している。例えば、Web ページ $\{A, B, C, D, E\}$ に対し、ユーザのアクティブセッションが $act = \langle A, B, C \rangle$ 、Large LCS 集合が、 $lcs_1 = \langle A, B, C, D, E \rangle$, $lcs_2 = \langle A, D, B, C \rangle$, $lcs_3 = \langle B, C, E \rangle$ であったとする。このとき、 lcs_1, lcs_2, lcs_3 から推薦候補となる Web ページは、 $\{D, E\}$ となる。ここでは、ページ D は一つの LCS、 lcs_1 から、ページ E は二つの LCS、 lcs_1, lcs_3 から、推薦候補となっているため、ページ E の推薦スコアの方がページ D の推薦スコアよりも高く算出される。

篠田ら [10] は、ユーザの行動履歴から行動パターンを抽出することで、ユーザに行動ナビゲーション情報を推薦するシステムを提案している。具体的には、複数のユーザ間の行動履歴を比較し、類似した行動履歴を持つ他のユーザの行動パターンを参照し、それに基づいて推薦を行っている。

石塚ら [94] は、ユーザの移動軌跡データの類似検索手法を提案している。観光地等において、多くのユーザの移動軌跡データと対象ユーザの移動軌跡データの類似検索を行うことにより、対象ユーザの移動軌跡に沿った観光施設等の情報を提供しようというものである。

以上のように、ユーザの行動パターンに沿った情報推薦技術について、いくつかの研究を取り上げたが、基本方式としては、蓄積されたユーザのアクセスログや行動履歴データなどから行動パターンを抽出し、直近のユーザの行動と類似する行動パターンに沿って推薦情報を決定している。人には日常的に何らかの行動パターンがあり、この行動パターンを考慮することで、ユーザにとって有益な情報推薦が可能となる。本論文において提案する情報推薦方式もユーザの行動パターンを考慮したものである。

ただし、本研究では「ユーザの行動パターンも推薦時コンテキストに応じて変化する」ことに着眼している。本章で提案する情報推薦方式では、まず、ユーザの行動履歴データベースから、そのユーザの一般的な行動パターンを抽出するのではなく、推薦時コンテキストに応じた限定的な行動パターンの抽出を行う。そして、この限定的な行動パターンに沿った情報推薦を行う。

6.3 提案方式の基本アイデア

6.3.1 概要

本節では、ユーザが現在おかれている状況およびユーザの時系列的な行動履歴をユーザコンテキストとして扱った、コンテキスト依存型情報推薦方式 (以降、提案方式) について説明する。

一般に、人は日常生活において、そのときおかれている状況に加え、その時点まで取ってきた行動や、その時点以降に予定している行動などに影響を受け、あらゆる意思決定を行っているといえる。例えば、

- (a) 「恋人といるとき、食事をした後は、バーで飲みたい」
- (b) 「一人にいるとき、買物を済ませた後は、カフェでゆっくり休みたい」

などといった場面では、“恋人といる” や “一人にいる” といった現在の状況に加え、“食事をした” や “買物を済ませた” といった、それまで取ってきた行動に影響を受け、それぞれ “バーで飲む”、“カフェで休む” といった行動を決定しているといえる。

(a) の例の場合、情報推薦システムでは、“恋人といる”、“食事をした” というユーザコンテキストを考慮して、近くのバーに関する情報を推薦することにより、ユーザを支援することができる。

提案方式では、第 3 章で述べた推薦時コンテキストおよび過去/未来コンテキストを考慮して、ユーザへの推薦情報を決定する。

6.3.2 問題設定

推薦時コンテキストと過去/未来コンテキスト

いま、推薦システムを利用するユーザ (以降、対象ユーザとよぶ) について考

える．対象ユーザの，推薦システムから推薦を受ける時点のコンテキストが，

$$C = \begin{cases} [\text{時刻}] = \text{“20:00”} \\ [\text{休日/非休日}] = \text{“休日”} \\ [\text{予算}] = \text{“5000円”} \\ [\text{同伴者}] = \text{“恋人”} \end{cases} \quad (25)$$

であったとする．また，対象ユーザはこの時点までに，

$$T = \begin{cases} \text{“買物する” at “17:00”} \\ \text{“食事する” at “18:00”} \end{cases} \quad (26)$$

という行動を取っていたとする．

このとき，対象ユーザは，これら C および T の両方に影響を受けて，情報選択に対する価値判断を行うものと推測することができる．第3章で述べた定義に則ると， C は推薦時コンテキスト， T は過去コンテキストであり，これらをユーザコンテキストとして扱う．提案方式は，これら C および T を総合的に考慮したうえで，ユーザへの推薦情報を決定するものである．

なお，本節では，推薦時コンテキストは，次のように，[時刻] や [予算] などの P 個のコンテキストパラメタの値 $c_p (1 \leq p \leq P)$ を各次元の要素とした P 次元特徴ベクトルで表現する．

$$C = (c_1, c_2, \dots, c_P) \quad (27)$$

また，過去コンテキストは，次のように表現する．

$$T = \langle (t_1, a_1), (t_2, a_2), \dots, (t_Q, a_Q) \rangle \quad (28)$$

ここで， $(t_q, a_q) (1 \leq q \leq Q)$ は，時刻 t_q においてユーザが取った行動 a_q を表す．つまり，時系列行動履歴は，時間情報を含む長さ Q の行動列 (または行動シーケンスとよぶ) として表現される．

さらに，未来コンテキストも過去コンテキストと同様に，式 (28) のように，行動シーケンスの形式で表現される．未来コンテキストとしての行動とは，対象ユーザが未来に予定している行動のことを表す．未来コンテキストの扱いについ

ては、基本的には過去コンテキストの扱いと同等であるため、本節では過去コンテキストを扱ったときの例で説明を行う。

行動履歴データベース

提案方式では、対象ユーザへの推薦情報を決定するために、他のユーザの行動パターンを参照する。ユーザの行動パターンは、推薦システムを利用するユーザー一人一人の行動履歴データベースから抽出される。

行動履歴データベースは、ユーザがこれまでに取った行動を、時系列に記録したものである。データの属性として、以下の情報をもつ。

- セッション ID (SID)
- 日時
- ユーザが取った行動
- 行動に関連して利用したアイテム
- 行動を取った時点のコンテキスト

行動履歴データベースの例を表 10 に示す。例えば、表 10 の第 1 番目のデータは、2008 年 12 月 1 日 12:00 に、ユーザが“食事する”ために、[和食]の店を利用したことを表す。特にこの時点のコンテキストは、

$$C = \begin{cases} [\text{時刻}] = \text{“12:00”} \\ [\text{休日/非休日}] = \text{“非休日”} \\ [\text{予算}] = \text{“3000 円”} \\ [\text{同伴者}] = \text{“なし”} \end{cases} \quad (29)$$

ということを示している。

ここで、ある期間においてユーザが取った一連の行動をまとめて 1 セッションとよぶ。1 セッションの期間は、推薦システムを設計する際に、そのシステムの目的に応じて設定する必要がある。表 10 では、1 日を 1 セッションの期間として

表 10 ユーザの行動履歴データベース (例)

SID	日時	行動	利用アイテム	コンテキスト			
				時刻	休日/非休日	予算(円)	同伴者
1	2008/12/1	12:00 食事する	和食	12:00	非休日	3,000	なし
		15:00 買物する	書店・古本	15:00	非休日	3,000	なし
2	2008/12/2	13:00 遊ぶ	ボウリング	13:00	休日	5,000	友人
		16:00 買物する	CD・DVD	16:00	休日	5,000	友人
		18:00 食事する	焼肉・韓国料理店	18:00	休日	5,000	友人
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

表 11 ユーザの行動シーケンスデータベース (例)

SID	行動シーケンス
1	<(12, "食事する", C_{11}), (15, "買物する", C_{12})>
2	<(13, "遊ぶ", C_{21}), (16, "買物する", C_{22}), (18, "食事する", C_{23})>
⋮	⋮

いる．セッション ID は，データベース内の全セッションに対して，通しで付けた番号である．第 k 番目のセッション ($SID = k$) は，次のようにシーケンス s_k として表記される．

$$s_k = \langle (t_1, a_1, C_1), (t_2, a_2, C_2), \dots \rangle \quad (30)$$

$(t_r, a_r, C_r) (r \geq 1)$ は，時刻 t_r ，コンテキスト C_r において，ユーザが取った行動 a_r を表す．つまり，第 k 番目のセッションは，時間情報およびコンテキスト情報を含む行動シーケンスとして表現される．行動履歴データベース Ω は，次式のように行動シーケンスの集合 S として表現することができる．

$$\Omega = S = \{s_1, s_2, \dots\} \quad (31)$$

行動履歴データベース Ω を，行動シーケンスデータベースの形式で表したものを，表 11 に示す．

また，行動についても，その表現方法は推薦システムの目的に応じて決定する必要がある．例えば，ユーザの日常的な行動を対象とした推薦システムの場合には，行動を {“食事する”，“酒を飲む”，“買物する”，“遊ぶ”，...} などと表現する．また，対象を飲食店に限定し，{“和食”，“居酒屋”，“カフェ”，“バー”，...} などのように，利用店のジャンルで表現することもできる．さらには，奈良の観光地における観光施設を対象とした推薦システムの場合には，{“東大寺”，“興福寺”，“春日大社”，...} などのような固有名詞での表現も可能である．

6.3.3 推薦手順

いま，推薦システムを利用する対象ユーザ A について考える．対象ユーザ A の推薦時コンテキストは C^A ，過去コンテキストは T^A であるとする．また，対象ユーザ A の他に，推薦システムを利用する $N(N \geq 1)$ 人のユーザ $U = \{u_1, u_2, \dots, u_N\}$ (以降，他ユーザとよぶ) が存在するとする．各ユーザ $u_i (i = 1, 2, \dots, N)$ には，各々に対し，それぞれ行動履歴データベース Ω^{u_i} が作成されている．このとき，以下の手順にしたがって対象ユーザ A への推薦情報を判定する．

- (1) 類似ユーザの選定
- (2) 推薦時コンテキスト C^A に関連する行動シーケンス集合の特定
- (3) 推薦時コンテキスト C^A における行動パターンの抽出
- (4) 過去コンテキスト T に合った行動パターンの特定
- (5) 特定された行動パターンに基づく行動予測スコアの算出
- (6) 行動予測スコアに基づく推薦候補アイテムのランキング

次節以降，各手順について説明する．

< 例 >

なお，本節では，各手順に沿って例を交えながら提案方式の説明を行う．説明用の例として，ここでは，

6.3.2 項「推薦時コンテキストと過去/未来コンテキスト」で例示した，式 (25) および式 (26) を，それぞれ C^A および T^A として用いる．ここでは， C^A は次の

ように 6 次元特徴ベクトルで表現する .

$$\begin{aligned}
 C^A &= \begin{cases} [\text{時刻}] = \text{“20 : 00”} \rightarrow c_1 = 0.833 \\ [\text{休日/非休日}] = \text{“休日”} \rightarrow c_2 = 1 \\ [\text{予算}] = \text{“5000 円”} \rightarrow c_3 = 0.5 \\ [\text{同伴者}] = \text{“恋人”} \rightarrow (c_4, c_5, c_6) = (0, 0, 1) \end{cases} & (32) \\
 &= (c_1, c_2, c_3, c_4, c_5, c_6) \\
 &= (0.833, 1, 0.5, 0, 0, 1)
 \end{aligned}$$

T^A は , 次のように表現する .

$$\begin{aligned}
 T^A &= \begin{cases} \text{“買物する” at “17 : 00”} \\ \text{“食事する” at “18 : 00”} \end{cases} & (33) \\
 &= \langle (t_1, a_1), (t_2, a_2) \rangle \\
 &= \langle (17, \text{“買物する”}), (18, \text{“食事する”}) \rangle
 \end{aligned}$$

なお , この例では , ユーザが繁華街を回遊するという場面を想定しており , 取りうる行動として , { “食事する” , “お茶する” , “バーで飲む” , “買物する” , “遊ぶ” } の五つの行動が定義されているものとする . また , ここでは , 他ユーザとしてユーザ $U = \{B, C, D, E\}$ が存在するものとする . 各ユーザの行動履歴データベースは , 図 32 に示したとおりのものであるとし , それぞれ , $\Omega^B, \Omega^C, \Omega^D, \Omega^E$ とする . 以降 , これらをもとに各手順について説明する .

類似ユーザの選定

提案方式では , 対象ユーザと嗜好の類似した他ユーザの情報をを用いて推薦を行う , 協調フィルタリング [4] を取り入れている . 協調フィルタリングとは , 例えば , ユーザ X がアイテム I を好むとすると , アイテム I を好む他のユーザ Y が好むアイテム J を見つけ出し , そのアイテム J をユーザ X に推薦するというものである . これは 「嗜好が類似するユーザは同じアイテムを好む」 という仮説に基づく推薦方式である .

(a) ユーザBの行動履歴データベース

SID	日時	行動	利用アイテム	コンテキスト				C ^A との関連度
				時刻	休日/非休日	予算(円)	同伴者	
1	2008/12/2	18:00 買物する	家電・パソコン	18:00	非休日	2,000	友人	0.343
		19:00 食事する	焼肉・韓国料理	19:00				
		20:00 遊ぶ	カラオケ	20:00				
2	2008/12/5	18:00 食事する	イタリア・フランス料理	18:00	休日	4,000	恋人	0.998
		19:00 パーで飲む	BAR	19:00				
3	2008/12/16	17:00 買物する	CD・DVD	17:00	非休日	3,000	友人	0.354
		18:00 食事する	焼肉・韓国料理	18:00				
		20:00 遊ぶ	カラオケ	20:00				
4	2008/12/20	16:00 お茶する	お菓子・ケーキ	16:00	休日	5,000	恋人	0.998
		17:00 買物する	洋服・和服	17:00				
		18:00 食事する	イタリア・フランス料理	18:00				
		20:00 遊ぶ	カラオケ	20:00				
5	2008/12/21	17:00 お茶する	お菓子・ケーキ	17:00	休日	7,000	恋人	0.993
		19:00 買物する	洋服・和服	19:00				
		20:00 食事する	中華	20:00				
		21:00 パーで飲む	BAR	21:00				

(b) ユーザCの行動履歴データベース

SID	日時	行動	利用アイテム	コンテキスト				C ^A との関連度
				時刻	休日/非休日	予算(円)	同伴者	
1	2008/12/3	11:00 食事する	和食	11:00	休日	1,000	なし	0.554
		12:00 お茶する	カフェ・喫茶	12:00				
2	2008/12/5	19:00 食事する	イタリア・フランス料理	19:00	休日	5,000	恋人	1.000
		20:00 パーで飲む	BAR	20:00				
3	2008/12/10	12:00 食事する	中華	12:00	休日	1,000	なし	0.576
		14:00 お茶する	カフェ・喫茶	14:00				
4	2008/12/15	19:00 食事する	イタリア・フランス料理	19:00	休日	6,000	恋人	0.998
		20:00 遊ぶ	カラオケ	20:00				
5	2008/12/17	11:00 食事する	中華	11:00	非休日	3,000	友人	0.289
		12:00 買物する	書店・古本	12:00				
		14:00 遊ぶ	ボウリング	14:00				

(c) ユーザDの行動履歴データベース

SID	日時	行動	利用アイテム	コンテキスト				C ^A との関連度
				時刻	休日/非休日	予算(円)	同伴者	
1	2008/12/2	18:00 遊ぶ	ピリヤード	18:00	休日	3,000	恋人	0.993
		19:00 パーで飲む	BAR	19:00				
2	2008/12/9	18:00 食事する	和食	18:00	非休日	3,000	友人	0.360
		19:00 お茶する	カフェ・喫茶	19:00				
		20:00 遊ぶ	カラオケ	20:00				
3	2008/12/16	19:00 食事する	イタリア・フランス料理	19:00	休日	4,000	恋人	0.998
		21:00 パーで飲む	BAR	21:00				
4	2008/12/18	17:00 買物する	洋服・和服	17:00	休日	7,000	恋人	0.993
		18:00 食事する	中華	18:00				
		19:00 遊ぶ	カラオケ	19:00				
		20:00 パーで飲む	BAR	20:00				
5	2008/12/20	12:00 食事する	中華	12:00	休日	1,000	なし	0.573
		13:00 お茶する	カフェ・喫茶	13:00				

(d) ユーザEの行動履歴データベース

SID	日時	行動	利用アイテム	コンテキスト				C ^A との関連度
				時刻	休日/非休日	予算(円)	同伴者	
1	2008/12/5	17:00 買物する	CD・DVD	17:00	非休日	2,000	友人	0.329
		18:00 食事する	和食	18:00				
2	2008/12/11	18:00 食事する	和食	18:00	休日	4,000	恋人	0.998
		19:00 遊ぶ	ボウリング	19:00				
3	2008/12/16	18:00 買物する	洋服・和服	18:00	非休日	3,000	友人	0.360
		19:00 食事する	中華	19:00				
		20:00 パーで飲む	BAR	20:00				
4	2008/12/19	18:00 お茶する	カフェ・喫茶	18:00	休日	1,000	なし	0.611
		19:00 食事する	中華	19:00				
5	2008/12/20	18:00 食事する	和食	18:00	休日	7,000	恋人	0.993
		19:00 遊ぶ	ボウリング	19:00				

図 32 ユーザ {B, C, D, E} の行動履歴データベース

この協調フィルタリングを適用するために、提案方式では、まず N 人の他ユーザ U の中から、対象ユーザ A と嗜好が類似する $M(1 \leq M \leq N)$ 人のユーザ $V = \{v_1, v_2, \dots, v_M\} (V \subseteq U)$ (以降、類似ユーザとよぶ) を選定する。

ここで、類似ユーザはユーザ類似度に基づいて判定される。対象ユーザ A と他ユーザ u_i のユーザ類似度は、次式で定義される。

$$\text{sim}^{user}(A, u_i) \quad (34)$$

すなわち、類似ユーザ V は、

$$\begin{aligned} V &= \{v_1, v_2, \dots, v_M\} \\ &= \{u_i | \text{sim}^{user}(A, u_i) \geq \text{threshold}^{user}\} \end{aligned} \quad (35)$$

となる。ここで、 threshold^{user} は、類似ユーザ選定のための閾値である。

なお、式 (34) の算出方法の定義については、提案方式においては制約はなく、アプリケーションに応じて決定する。

< 例 >

本節の例では、図 33 のように、推薦時コンテキスト C^A における、各ユーザの施設ジャンルの利用割合に基づいて、ユーザ類似度を求める。例えば、図 33(a) では、 C^A においては、“食事する” という行動に関して、ユーザ A は [中華] および [イタリア・フランス料理] をそれぞれ 50% の割合で利用していることを示している。同様に、他ユーザの利用施設ジャンルの割合を調べたとき、この割合の類似度に基づいて、ユーザ A との類似度を算出する。最終的には、各行動に関する類似度の平均から、ユーザ類似度が算出される。なお、この例では、 $\text{threshold}^{user} = 0.8$ とする。故に、図 33(f) より、 $V = \{B, C, D\}$ を対象ユーザ A の類似ユーザとして扱う。

推薦時コンテキストに関連する行動シーケンス集合の特定

選定した類似ユーザ V の行動履歴データベース Ω^V から、対象ユーザ A の推薦時コンテキスト C^A に関連する行動シーケンスの集合 S^* を特定する。

(a) "食事する"に関するユーザ類似度					
	利用施設ジャンルの割合				ユーザAとの類似度
	和食	中華	焼肉・韓国料理	イタリア・フランス料理	
ユーザA	0.0%	50.0%	0.0%	50.0%	1.000
ユーザB	0.0%	33.3%	0.0%	66.7%	0.849
ユーザC	0.0%	0.0%	0.0%	100.0%	0.707
ユーザD	0.0%	50.0%	0.0%	50.0%	1.000
ユーザE	100.0%	0.0%	0.0%	0.0%	0.000

(b) "お茶する"に関するユーザ類似度			
	利用施設ジャンルの割合		ユーザAとの類似度
	カフェ・喫茶	お菓子・ケーキ	
ユーザA	25.0%	75.0%	1.000
ユーザB	0.0%	100.0%	0.849
ユーザC	-	-	-
ユーザD	-	-	-
ユーザE	-	-	-

(c) "バーで飲む"に関するユーザ類似度		
	利用施設ジャンルの割合	ユーザAとの類似度
	BAR	
ユーザA	100.0%	1.000
ユーザB	100.0%	1.000
ユーザC	100.0%	1.000
ユーザD	100.0%	1.000
ユーザE	-	-

(d) "買物する"に関するユーザ類似度					
	利用施設ジャンルの割合				ユーザAとの類似度
	家電・パソコン	書店・古本	CD・DVD	洋服・和服	
ユーザA	0.0%	0.0%	0.0%	100.0%	1.000
ユーザB	0.0%	0.0%	0.0%	100.0%	1.000
ユーザC	-	-	-	-	-
ユーザD	0.0%	0.0%	0.0%	100.0%	1.000
ユーザE	-	-	-	-	-

(e) "遊ぶ"に関するユーザ類似度					
	利用施設ジャンルの割合				ユーザAとの類似度
	カラオケ	ゲームセンター	ボウリング	ビリヤード	
ユーザA	75.0%	0.0%	0.0%	25.0%	1.000
ユーザB	100.0%	0.0%	0.0%	0.0%	0.849
ユーザC	100.0%	0.0%	0.0%	0.0%	0.849
ユーザD	50.0%	0.0%	0.0%	50.0%	0.884
ユーザE	0.0%	0.0%	100.0%	0.0%	0.000

(f) ユーザAとのユーザ類似度	
ユーザ	類似度
ユーザB	0.969
ユーザC	0.885
ユーザD	0.974
ユーザE	0.000



図 33 ユーザ $\{B, C, D, E\}$ のユーザ A とのユーザ類似度

いま，類似ユーザ $v_j (j = 1, 2, \dots, M)$ の行動履歴データベース Ω^{v_j} 内の第 k 番目のセッションを，行動シーケンス $s_k^{v_j}$ として，

$$s_k^{v_j} = \langle (t_1^{v_j}, a_1^{v_j}, C_1^{v_j}), (t_2^{v_j}, a_2^{v_j}, C_2^{v_j}), \dots \rangle \quad (36)$$

とする．このとき， C^A と $s_k^{v_j}$ との関連度 $\text{rel}(C^A, s_k^{v_j})$ は，次式で求められる．

$$\text{rel}(C^A, s_k^{v_j}) = \frac{1}{l} \sum_{h=1}^l \text{sim}^{\text{context}}(C^A, C_h^{v_j}) \quad (37)$$

ここで， $\text{sim}^{\text{context}}(C^A, C_h^{v_j})$ は，コンテキスト C^A とコンテキスト $C_h^{v_j}$ との類似度であり，コンテキスト類似度とよぶ．コンテキスト類似度は一般的なベクトル間の類似度として求められる．一般によく用いられているコサイン尺度 [5] を用いると，コンテキスト C_1 とコンテキスト C_2 とのコンテキスト類似度は，

$$\text{sim}^{\text{context}}(C_1, C_2) = \frac{C_1 \cdot C_2}{\|C_1\| \|C_2\|} \quad (38)$$

により求められる．

このとき， S^* は，

$$\begin{aligned} S^* &= \{s_1^*, s_2^*, \dots\} \\ &= \{s_k^{v_j} | \text{rel}(C^A, s_k^{v_j}) \geq \text{threshold}^{\text{context}}\} \end{aligned} \quad (39)$$

となる．ここで， $threshold^{context}$ は，推薦時コンテキスト C^A と関連するセッションを選定するための閾値である．また， s_k^* は，

$$s_k^* = \langle (t_1^*, a_1^*), (t_2^*, a_2^*), \dots \rangle \quad (40)$$

となる．

なお，この S^* により構成される行動履歴データベースを

$$\Omega^* = S^* \quad (41)$$

とする．

< 例 >

類似ユーザ $V = \{B, C, D\}$ の各行動履歴データベース中における各行動シーケンスと推薦時コンテキスト C^A との関連度を求めると，図 32 中の「 C^A との関連度」に示すとおりとなる．

いま， $threshold^{context} = 0.8$ とすると，類似ユーザ B の行動履歴データベース Ω^B から $\{s_2, s_4, s_5\}$ ，類似ユーザ C の行動履歴データベース Ω^C から $\{s_2, s_4\}$ ，類似ユーザ D の行動履歴データベース Ω^D から $\{s_1, s_3, s_4\}$ ，の行動シーケンス集合（図 32 中のグレー部分）が， C^A に関連する行動シーケンス集合として特定される．

最終的に，これら特定されたセッションを統合し，表 12 のような新たな行動履歴データベース Ω^* を得る．これを行動シーケンスデータベースの形式で表現すると，表 13 のようになる．

推薦時コンテキストにおける行動パターンの抽出

特定された行動履歴データベース Ω^* から，類似ユーザの行動パターンを抽出する．

基本的な方法として，各行動シーケンスごとに行動シーケンスデータベース内における出現頻度を調べ，事前に設定された最小頻度を超える頻度を持つ行動シーケンスを，行動パターンとする．

表 12 推薦時コンテキスト C^A に関連した行動履歴データベース

SID	日時	行動
1	2008/12/5	18:00 食事する
		19:00 バーで飲む
2	2008/12/20	16:00 お茶する
		17:00 買物する
		18:00 食事する
		20:00 遊ぶ
3	2008/12/21	17:00 お茶する
		19:00 買物する
		20:00 食事する
		21:00 バーで飲む
4	2008/12/5	19:00 食事する
		20:00 バーで飲む
5	2008/12/15	19:00 食事する
		20:00 遊ぶ
6	2008/12/2	18:00 遊ぶ
		19:00 バーで飲む
7	2008/12/16	19:00 食事する
		21:00 バーで飲む
8	2008/12/18	17:00 買物する
		18:00 食事する
		19:00 遊ぶ
		20:00 バーで飲む

表 13 推薦時コンテキスト C^A に関連した行動シーケンスデータベース

SID	行動シーケンス
1	<(18, "食事する"), (19, "バーで飲む")>
2	<(16, "お茶する"), (17, "買物する"), (18, "食事する"), (20, "遊ぶ")>
3	<(17, "お茶する"), (19, "買物する"), (20, "食事する"), (21, "バーで飲む")>
4	<(19, "食事する"), (20, "バーで飲む")>
5	<(19, "食事する"), (20, "遊ぶ")>
6	<(18, "遊ぶ"), (19, "バーで飲む")>
7	<(19, "食事する"), (21, "バーで飲む")>
8	<(17, "買物する"), (18, "食事する"), (19, "遊ぶ"), (20, "バーで飲む")>

なお，行動履歴データベースからの行動パターンの抽出に関しては，6.2.1 項で取り上げたような既存手法を適用することにより，容易に行うことが可能である．6.4 節では，既存の行動パターン抽出手法を適用したときの例を説明する．

抽出された推薦時コンテキスト C^A における行動パターンの集合 F^* を次のように表記する．

$$F^* = \{f_1^*, f_2^*, \dots\} \quad (42)$$

特に，

$$f_g^* = \langle (t_1, a_1), (t_2, a_2), \dots \rangle \quad (43)$$

と表す．

< 例 >

表 14 各行動シーケンスの出現頻度と行動パターン

行動シーケンス	出現頻度	行動パターン
<"食事する", "バーで飲む">	5	f_1^*
<"食事する", "遊ぶ">	3	f_2^*
<"お茶する", "食事する">	2	f_3^*
<"お茶する", "バーで飲む">	1	
<"お茶する", "買物する">	2	f_4^*
<"お茶する", "遊ぶ">	1	
<"買物する", "食事する">	3	f_5^*
<"買物する", "バーで飲む">	2	f_6^*
<"買物する", "遊ぶ">	2	f_7^*
<"遊ぶ", "バーで飲む">	2	f_8^*
<"食事する", "遊ぶ", "バーで飲む">	1	
<"お茶する", "食事する", "バーで飲む">	1	
<"お茶する", "食事する", "遊ぶ">	1	
<"お茶する", "買物する", "食事する">	2	f_9^*
<"お茶する", "買物する", "バーで飲む">	1	
<"お茶する", "買物する", "遊ぶ">	1	
<"買物する", "食事する", "バーで飲む">	2	f_{10}^*
<"買物する", "食事する", "遊ぶ">	2	f_{11}^*
<"買物する", "遊ぶ", "バーで飲む">	1	
<"お茶する", "買物する", "食事する", "バーで飲む">	1	
<"お茶する", "買物する", "食事する", "遊ぶ">	1	
<"買物する", "食事する", "遊ぶ", "バーで飲む">	1	

Ω^* に対し，各行動シーケンスの出現頻度を調べたものを，表 14 に示す．ここで，最小頻度を 2 とすると，表 14 に示す，

$$F^* = \{f_1^*, f_2^*, \dots, f_{11}^*\} \quad (44)$$

の 11 個の行動シーケンス (表 14 中のグレー部分) が行動パターンとして抽出される．

過去コンテキストに合った行動パターンの特定

抽出された行動パターン集合 F^* の中から，対象ユーザ A の過去コンテキスト T^A に合った行動パターン f_g^* を特定する．

いま，

$$T^A = \langle (t_1, a_1), (t_2, a_2) \rangle \quad (45)$$

とする．このとき， T^A の部分行動シーケンスとして，時間情報を取り除いた以

下のいずれかを含む行動パターン f_g^* が特定される .

$$\langle a_1, x \rangle \quad (46)$$

$$\langle a_2, x \rangle \quad (47)$$

$$\langle (a_1, a_2), x \rangle \quad (48)$$

ただし , x は長さ任意の行動シーケンスである . また , (a_1, a_2) は , a_1 と a_2 の出現の順序は問わず , a_1 と a_2 の両方が出現することを表す .

ここで特定された行動パターン集合を

$$F^{**} = \{f_1^{**}, f_2^{**}, \dots\} \quad (49)$$

とする .

なお , ここでは過去コンテキストを扱った場合を説明したが , 未来コンテキストを扱った場合も同様である . いま , 未来コンテキストとして ,

$$T_2^A = \langle (t_4, a_4), (t_5, a_5) \rangle \quad (50)$$

があったとする . このとき , 式 (46)(47)(48) と合わせて ,

$$\langle x, a_4 \rangle \quad (51)$$

$$\langle x, a_5 \rangle \quad (52)$$

$$\langle x, (a_4, a_5) \rangle \quad (53)$$

のいずれかを含む行動パターンを特定する .

< 例 >

対象ユーザ A の過去コンテキスト T^A は , 式 (33) であるから , 表 14 に示す行動パターン集合 F^* の中から ,

$$\langle \text{“買物する”}, x \rangle \quad (54)$$

$$\langle \text{“食事する”}, x \rangle \quad (55)$$

$$\langle (\text{“買物する”}, \text{“食事する”}), x \rangle \quad (56)$$

表 15 過去コンテキスト T^A に合った頻出行動パターン

行動シーケンス	出現頻度	行動パターン
<"食事する", "バーで飲む">	5	f_1^{**}
<"食事する", "遊ぶ">	3	f_2^{**}
<"買物する", "食事する">	3	f_3^{**}
<"買物する", "バーで飲む">	2	f_4^{**}
<"買物する", "遊ぶ">	2	f_5^{**}
<"買物する", "食事する", "バーで飲む">	2	f_6^{**}
<"買物する", "食事する", "遊ぶ">	2	f_7^{**}

のいずれかを含む行動パターン集合が、過去コンテキスト T^A に合った行動パターン集合 F^{**} として特定される。本節の例では、表 15 に示す、

$$F^{**} = \{f_1^{**}, f_2^{**}, \dots, f_7^{**}\} \quad (57)$$

の 7 個の行動パターンが特定される。

特定された行動パターンに基づく行動予測スコアの算出

特定された行動パターン集合 F^{**} に基づいて、各行動の予測スコアを算出する。ここで、行動 x の予測スコアとは、対象ユーザ A が次に行動 x を取ることに對する確信度であり、このスコアが高いほど、ユーザが確信的に次に行動 x を取ることを意味する。

基本的には、行動パターン集合 F^{**} の出現頻度に基づいて、各行動の予測スコアを算出する。具体的には、行動 x の予測スコアは、

$$\langle a_1, x \rangle \quad (58)$$

$$\langle a_2, x \rangle \quad (59)$$

$$\langle (a_1, a_2), x \rangle \quad (60)$$

の各行動シーケンスの出現頻度の和により求められる。

ただし、この行動予測スコアの算出方法は、本項「推薦時コンテキストにおける行動パターンの抽出」において、適用する行動パターン抽出手法に依存する。各手法を適用したときの行動予測スコアの算出方法については、6.4 節にて述べる。

< 例 >

表 16 各行動の予測スコア

行動	推薦スコア
"食事する"	3
"お茶する"	0
"バーで飲む"	9
"買物する"	0
"遊ぶ"	7

行動シーケンス	出現頻度	頻出行動パターン
<"食事する", "バーで飲む">	5	f_1^{**}
<"食事する", "遊ぶ">	3	f_2^{**}
<"買物する", "食事する">	3	f_3^{**}
<"買物する", "バーで飲む">	2	f_4^{**}
<"買物する", "遊ぶ">	2	f_5^{**}
<"買物する", "食事する", "バーで飲む">	2	f_6^{**}
<"買物する", "食事する", "遊ぶ">	2	f_7^{**}



 "バーで飲む"に対応するスコア
 $= 5 + 2 + 2 = 9$

図 34 行動の予測スコアの算出例 (行動 "バーで飲む" の場合)

表 15 に基づいて、各行動 { "食事する", "お茶する", "バーで飲む", "買物する", "遊ぶ" } の予測スコアを算出すると、表 16 に示すようになる。図 34 は、行動 "バーで飲む" の予測スコアの算出例を示したものである。行動 "バーで飲む" を含む行動パターンは、 $\{f_1^{**}, f_4^{**}, f_6^{**}\}$ であるため、これらの出現頻度の和が行動 "バーで飲む" の予測スコアとなる。ここで、行動 "お茶する" および "買物する" に関しては、これらの行動を含む行動パターンが発見されていないため、行動予測スコアは 0 となる。

行動予測スコアに基づく推薦候補アイテムのランキング

最終的に行動予測スコアに基づき、ユーザへの推薦アイテムを選定する。

なお、第 4 章では、膨大なアイテム集合の中から、ユーザの嗜好およびユーザの推薦時コンテキストに合ったアイテム集合を、推薦候補アイテム集合として選定する推薦方式を提案している。本章においても、最初の段階としての推薦候補アイテム集合の選定に関しては、第 4 章の推薦方式により行う。本章での提案方

式は、算出された各行動予測スコアに基づき、あらかじめ選定された推薦候補アイテム集合に対し、ランキングを行うものである。

なお、推薦候補となるアイテムには、6.3.2項「行動履歴データベース」で説明したような、ユーザの行動が関連付けられているものとし、高い予測スコアを持つ行動に関連付けられたアイテムがランキング上位に提示される。

< 例 >

本節の例では、行動“バーで飲む”の行動予測スコアが最も高くなったため、推薦候補アイテム集合の中で、行動“バーで飲む”に関連付けられたアイテム(この場合では「バー」)が、次いで、行動“遊ぶ”に関連付けられたアイテム(例えば、「ボウリング」や「カラオケ」)が、ランキング上位に提示されることになる。

6.4 行動パターン抽出手法の適用

本節では、提案方式における行動パターン抽出に、6.2.1項にて取り上げた、各行動パターン抽出手法を適用した例を示す。

6.4.1 シーケンシャルパターンマイニング手法の適用例

本項では、シーケンシャルパターンマイニング手法 [84] (以降、SPM と表記する) を適用した例を示す。

表 17 は、表 13 の行動シーケンスデータベースに対し、SPM を適用した結果を示している。ここで、表 17 中の count は出現回数、sup. はサポート値 (支持度)、conf. はコンフィデンス値 (確信度) を表す。それぞれ次のように算出される。

シーケンス $\langle a, b \rangle$ のサポート値 $\text{sup}(\langle a, b \rangle)$ は、シーケンスデータベース内におけるシーケンス $\langle a, b \rangle$ の出現頻度を意味し、次式で算出される。

$$\text{sup}(\langle a, b \rangle) = \frac{n(\langle a, b \rangle)}{N} \quad (61)$$

ここで、 $n(\langle a, b \rangle)$ はシーケンスデータベース内におけるシーケンス $\langle a, b \rangle$ を含むシーケンス数、 N はシーケンスデータベース内の全シーケンス数を表す。

表 17 シーケンシャルパターンマイニング手法による行動パターンの抽出結果
(minsup = 0.2)

行動シーケンス	count	sup.	conf.
<"食事する","バーで飲む">	5	0.625	0.714
<"食事する","遊ぶ">	3	0.375	0.429
<"お茶する","食事する">	2	0.250	1.000
<"お茶する","買物する">	2	0.250	1.000
<"買物する","食事する">	3	0.375	1.000
<"買物する","バーで飲む">	2	0.250	0.667
<"買物する","遊ぶ">	2	0.250	0.667
<"遊ぶ","バーで飲む">	2	0.250	0.500
<"お茶する","買物する","食事する">	2	0.250	1.000
<"買物する","食事する","バーで飲む">	2	0.250	0.667
<"買物する","食事する","遊ぶ">	2	0.250	0.667

また、シーケンス $\langle a, b \rangle$ のコンフィデンス値 $\text{conf}(\langle a, b \rangle)$ は、シーケンスデータベース内において、アイテム a が発生した後にアイテム b が発生する割合を意味し、次式で算出される。

$$\text{conf}(\langle a, b \rangle) = \frac{\text{sup}(\langle a, b \rangle)}{\text{sup}(\langle a \rangle)} \quad (62)$$

$\text{conf}(\langle a, b \rangle)$ は、すなわち、アイテム a が発生したという条件下において、アイテム b が発生する確信度を表す。

SPM では、サポート値が事前に設定された最小サポート値 minsup を満たすシーケンスが頻出パターンとして抽出される。表 17 の例では、minsup = 0.2 としたときの抽出結果を示している。

つづいて、6.3.3 項「過去コンテキストに合った行動パターンの特定」にしたがって過去コンテキスト T^A に合った行動パターンを特定すると、表 18 のようになる。

さらに、6.3.3 項「特定された行動パターンに基づく行動予測スコアの算出」にしたがって、表 18 から、各行動の予測スコアを算出する。なお、ここでは、確信度を表すコンフィデンス値に基づきスコアを算出することにする。例えば、行動“バーで飲む”の場合、予測スコアは、 $0.714 + 0.667 + 0.667 = 2.048$ となる。各行動の予測スコアの算出結果を表 19 に示す。

結果、推薦候補アイテム集合の中から、行動“バーで飲む”に関連付けられたアイテム、次いで、行動“遊ぶ”に関連付けられたアイテムが、ランキング上位に提示されることになる。

表 18 シーケンシャルパターンマイニング手法による過去コンテキストに合った行動パターンの特定結果

行動シーケンス	count	sup.	conf.
<"食事する","バーで飲む">	5	0.625	0.714
<"食事する","遊ぶ">	3	0.375	0.429
<"買物する","食事する">	3	0.375	1.000
<"買物する","バーで飲む">	2	0.250	0.667
<"買物する","遊ぶ">	2	0.250	0.667
<"買物する","食事する","バーで飲む">	2	0.250	0.667
<"買物する","食事する","遊ぶ">	2	0.250	0.667

表 19 シーケンシャルパターンマイニング手法による各行動の予測スコア

行動	行動予測スコア
"食事する"	1.000
"お茶する"	0.000
"バーで飲む"	2.048
"買物する"	0.000
"遊ぶ"	1.763

ここで、SPM を適用する際、最小サポート値 minsup の決め方が一つの検討事項となる。 minsup を大きくすることで、抽出できるパターン数は増加するが、これに伴い処理時間も増加する。したがって、抽出したいパターン数と処理時間の兼ね合いを考慮して、 minsup を決める必要がある。

また、本研究で扱うような情報推薦の分野においては、推薦に対する正確性と意外性の観点から次のことがいえると考えられる。 minsup を大きく設定することにより、より多くのユーザにとって主流な行動パターンが抽出される。したがって、推薦結果は多くのユーザに支持されるような正確なものになると考えられる。一方で、 minsup を小さく設定することにより、一部のユーザによる局所的な行動パターンも抽出される。このため、推薦結果の中には意外なものも含まれ、ユーザにとって新たな発見につながることも期待できる。この minsup は、目的とするアプリケーションに応じて決定する必要がある。できるだけ正確性を重視した推薦を行いたい場合には、 minsup を大きく設定し、推薦結果に発見性を求めるような場合には、 minsup を小さく設定する。

6.4.2 LCS 抽出手法の適用例

本項では、LCS 抽出手法 [90] [91] を適用した例を示す。

表 20 LCS 抽出手法による行動パターンの抽出結果 ($min.Count = 1, min.Length = 2$)

行動シーケンス	count
<“食事する”, “バーで飲む”>	9
<“お茶する”, “買物する”, “食事する”>	1
<“食事する”, “遊ぶ”>	2
<“買物する”, “食事する”, “遊ぶ”>	1
<“買物する”, “食事する”, “バーで飲む”>	1
<“遊ぶ”, “バーで飲む”>	1

表 21 LCS 抽出手法による過去コンテキストに合った行動パターンの特定結果

行動シーケンス	count
<“食事する”, “バーで飲む”>	9
<“食事する”, “遊ぶ”>	2
<“買物する”, “食事する”, “遊ぶ”>	1
<“買物する”, “食事する”, “バーで飲む”>	1

表 20 は、表 13 の行動シーケンスデータベースに対し、LCS 抽出手法を適用した結果を示している。ここで、表 20 中の count は出現回数を表す。

LCS 抽出手法では、count が $min.Count$ 以上であり、かつ長さが $min.Length$ 以上である LCS の集合を Large LCS 集合として扱い、これを抽出する [11]。表 20 の例では、 $min.Count = 1, min.Length = 2$ としたときの抽出結果を示している。

つづいて、6.3.3 項「過去コンテキストに合った行動パターンの特定」にしたがって過去コンテキスト T^A に合った行動パターンを特定すると、表 21 のようになる。

さらに、6.3.3 項「特定された行動パターンに基づく行動予測スコアの算出」にしたがって、表 21 から、各行動の予測スコアを算出する。なお、ここでは、count に基づきスコアを算出することにする。例えば、行動“バーで飲む”の場合、対応するスコアは、 $9 + 1 = 10$ となる。各行動の予測スコアの算出結果を表 22 に示す。

結果、推薦候補アイテム集合の中から、行動“バーで飲む”に関連付けられたアイテム、次いで、行動“遊ぶ”に関連付けられたアイテムが、ランキング上位に提示されることになる。

ここで、LCS 抽出手法を適用する際、 $min.Count$ および $min.Length$ の決め方が一つの検討事項となる。6.4.1 項で述べたことと同様、 $min.Count$ および $min.Length$ は、目的とするアプリケーションに応じて決定する必要がある。

表 22 LCS 抽出手法による各行動の予測スコア

行動	行動予測スコア
"食事する"	0
"お茶する"	0
"バーで飲む"	10
"買物する"	0
"遊ぶ"	3

6.5 評価実験

提案方式の有効性を評価するために被験者実験を行った。本実験では、日常的に繁華街において、飲食店や娯楽施設等を利用する場面を想定した、被験者の行動履歴データを作成し、これに基づいた推薦に対する評価を行う。

6.5.1 実験条件

本実験では、一般向け地域情報検索サイトである『ドコイク?』[95]に登録されている施設データを利用した。『ドコイク?』には、飲食店からショッピング施設、娯楽施設など、さまざまなジャンルの施設データが1000万件以上登録されている。

『ドコイク?』には、日本全国の施設データが登録されているが、本実験では「なんばエリア」を対象とした施設データを用いた。これは、「なんばエリア」が、本実験の対象とする被験者にとって身近な繁華街であるため、実験を行う上で、被験者が日常的に「なんばエリア」の施設を利用する場面を想定しやすくするためであり、かつそのときのユーザコンテキストを想像しやすくするためである。

『ドコイク?』に登録されている各種施設データのうち、本実験では、日常性、娯楽性のある施設データとして、表 23 に示す施設ジャンルに属するデータを選んだ。なお、表 23 に示す施設ジャンル名については、『ドコイク?』で使用されている表記をそのまま用いた。施設カテゴリについては、『ドコイク?』で用いられているカテゴリ分類を参考にしながら、表 23 に示す五つのカテゴリを設けた。さらに、各カテゴリに対応した行動として表 23 に示す行動を関連付けている。つまり、被験者は施設データを利用することで、関連する行動を取るということを意味する。本実験で用いた各施設カテゴリの登録データ件数は、表 23 に示すと

表 23 実験で用いた施設データ

関連する行動	施設カテゴリ	施設ジャンル	登録データ件数
食事する	飲食店	[和食], [中華], [ラーメン], [焼肉・韓国料理], [イタリア・フランス料理], [インド料理・カレー], [各国料理], [居酒屋]	1618
お茶する	軽食店	[カフェ・喫茶], [お菓子・ケーキ]	343
バーで飲む	バー	[BAR]	95
買物する	ショッピング施設	[家電・パソコン], [書店・古本], [CD・DVD], [洋服・和服]	587
遊ぶ	娯楽施設	[カラオケ], [ゲームセンター], [ボウリング], [ビリヤード]	72

おりである (2008 年 11 月 2 日現在)。

本実験における被験者は、22 歳から 28 歳の男女 8 名である。いずれの被験者も、「なんばエリア」には何度か訪れたことがあり、このエリアの雰囲気などについては十分把握している。

6.5.2 実験方法

本実験は、以下のフェーズにしたがって遂行する。

フェーズ 1：被験者の行動履歴データベースの作成

フェーズ 2：被験者間のユーザ類似度の算出

フェーズ 3：比較評価用の行動履歴データベースの特定

フェーズ 4：比較評価用の行動履歴データベースに基づく行動予測スコアの算出

フェーズ 5：スコア算出結果に対する評価

以下、各フェーズについて説明する。

フェーズ 1：被験者の行動履歴データベースの作成

まず、以下の手順にしたがって、あらかじめ各被験者の行動履歴データベースを作成する。

- (1) 被験者に、提示したコンテキスト (推薦時コンテキスト+過去コンテキスト) を想定してもらい、提示するコンテキストは、表 24 に示す推薦時コン

表 24 実験で用いた推薦時コンテキストおよび過去コンテキスト

コンテキストタイプ	段階	設定数	設定内容
推薦時 コンテキスト	-	6	「平日の20時に一人で」、「休日の20時に一人で」、 「平日の20時に友達と」、「休日の20時に友達と」、 「平日の20時に恋人と」、「休日の20時に恋人と」
過去 コンテキスト	1段階	3	「(19時, “食事する”)」、 「(19時, “バーで飲む”)」、 「(19時, “遊ぶ”)」
	2段階	3	「(18時, “食事する”), (19時, “バーで飲む”)」、 「(18時, “遊ぶ”), (19時, “バーで飲む”)」、 「(18時, “遊ぶ”), (19時, “食事する”)」
	3段階	1	「(17時, “遊ぶ”), (18時, “食事する”), (19時, “バーで飲む”)」

テキストおよび過去コンテキストの組合せ (例 ; 「休日の 20 時に一人で」 + 「(19 時, “食事する”)」) の中から選択する . ここで , 推薦時コンテキストに関しては , 被験者一人当たり , 6 個の設定コンテキストのうち 3 個を選ぶことにする . したがって , 被験者一人当たりの想定するコンテキスト数は , $3 \times (3 + 3 + 1) = 21$ 組となる .

- (2) 想定した各コンテキストにおいて , 『ドコイク?』に登録されている施設データの中から 20 件を (各施設カテゴリから 4 件ずつランダムに) 選び , 被験者に提示する . したがって , 被験者一人当たりの提示施設データ数は , $20 \times 21 = 420$ データとなる .
- (3) 各コンテキストにおいて提示した施設データに対し , 被験者に利用したい程度を 5 段階 {5:利用したい, 4:やや利用したい, 3:どちらでもない, 2:やや利用したくない, 1:利用したくない} で評価を行ってもらう .
- (4) 被験者評価に基づいて , 各被験者の行動履歴データを作成する . ここでは , 評価 4 以上とした施設データに関連する行動を対象に , 行動履歴データベースを作成する .

本実験での被験者は , $\{A, B, C, D, E, F, G, H\}$ の 8 名であり , うち , 被験者 A による実験の例を図 35 に示す . 図 35(a) は , 被験者 A への設問例であり , 図 35(b) のように , 被験者 A は , 提示されたコンテキストにおいて , 各施設データに対し , 5 段階の評価を行っている . この評価を 21 組のコンテキストにおいて行ってもらった . また , この被験者 A の評価に基づいて , 図 35(c) の行動履歴データベースが作成される .

表 25 各被験者の推薦時コンテキストの設定

推薦時コンテキスト	被験者							
	A	B	C	D	E	F	G	H
「平日の20時に一人で」	○	○	○	○				
「休日の20時に一人で」					○	○	○	○
「平日の20時に友達と」	○	○			○	○		
「休日の20時に友達と」			○	○			○	○
「平日の20時に恋人と」	○		○		○		○	
「休日の20時に恋人と」		○		○		○		○

なお、各被験者の推薦時コンテキストは、表 25 に示すとおりに設定した。例えば、被験者 A は、推薦時コンテキストとして、「平日の 20 時に一人で」、「平日の 20 時に友達と」、「平日の 20 時に恋人と」の 3 個を設定した。このように全被験者において設定コンテキストがばらけるようにした。

フェーズ 2：被験者間のユーザ類似度の算出

本実験では、以下の 3 タイプの被験者間のユーザ類似度を求める。

- (a) sim_a^{user} : コンテキストを考慮しないユーザ類似度
- (b) sim_b^{user} : 推薦時コンテキストのみを考慮したユーザ類似度
- (c) sim_c^{user} : 推薦時コンテキストおよび過去コンテキストの両方を考慮したユーザ類似度

基本的には、被験者間のユーザ類似度は、各施設ジャンルに対する評価値の相関により求められる。したがって、被験者 A と被験者 B のユーザ類似度 $\text{sim}^{user}(A, B)$ は、次式により表される。

$$\text{sim}^{user}(A, B) = \frac{\sum(e_{Ai} - \bar{e}_A)(e_{Bi} - \bar{e}_B)}{\sqrt{\sum(e_{Ai} - \bar{e}_A)^2} \sqrt{\sum(e_{Bi} - \bar{e}_B)^2}} \quad (63)$$

ここで、 e_{ji} は被験者 j の施設ジャンル i に与えた評価値、 \bar{e}_j は被験者 j が与えた平均評価値を表す。

このとき、 $\text{sim}_a^{user}(A, B)$ は、被験者 A および被験者 B それぞれの全評価データを対象に算出される。 $\text{sim}_b^{user}(A, B)$ は、被験者 A および被験者 B それぞれの評価データのうち、推薦時コンテキストが一致するデータを対象に算出され、 $\text{sim}_c^{user}(A, B)$

(a) 被験者Aへの設問

<p>あなたは今、休日の20時に一人で大阪のなんばにいます。 この日の、19時に飲食店で食しました。 このシチュエーションにおいて、あなたが利用したい施設を教えてください。 提示された次の各施設に対して、5段階評価で教えてください。</p>	<p>5: 利用したい 4: やや利用したい 3: どちらでもない 2: やや利用したくない 1: 利用したくない</p>
---	---

(b) 被験者Aの施設データに対する評価

行動	利用施設ジャンル	利用施設名	評価
食事する	中華 居酒屋	餃子の王将 日本橋でんどんタウン店	5-4-3-2- 1
食事する		甘太郎なんば店	5-4-3-2- 1
お茶する	お菓子・ケーキ カフェ・喫茶	もぐらやナンバ店	5- 4 -3-2-1
お茶する		カフェレスト316	5 -4-3-2-1
バーで飲む	BAR BAR	Girl's Barメルローズ	5-4-3-2- 1
バーで飲む		Stylish Bar midori	5-4-3-2- 1
買物する	洋服・和服 書店・古本	アンテルリネエール	5-4-3- 2 -1
買物する		波屋書房	5 -4-3-2-1
遊ぶ	ゲームセンター カラオケ	コスモナンバ店CDゲーム館	5-4-3-2- 1
遊ぶ		サウンド・バード	5-4-3-2- 1

(c) 被験者Aの行動履歴データベース

SID	日時	行動	利用施設 ジャンル	コンテキスト		
				時刻	休日/非休日	同伴者
1	2008/12/1	19:00 食事する	お菓子・ケーキ	19:00	休日	一人
		20:00 お茶する		20:00		
2	2008/12/2	19:00 食事する	カフェ・喫茶	19:00	休日	一人
		20:00 お茶する		20:00		
5	2008/12/5	19:00 食事する	書店・古本	19:00	休日	一人
		20:00 買物する		20:00		

図 35 被験者 A による実験の例

は、推薦時コンテキストおよび過去コンテキストの両方が一致するデータを対象に算出される。

本実験における被験者間の各タイプのユーザ類似度を、図 36 に示す。ここで、図 36 中の “NaN” は、被験者間で共通のコンテキストにおける評価データが存在しないため、ユーザ類似度が算出されていないことを表す。例えば、被験者 A と被験者 H では、表 25 より、共通となるコンテキストが設定されていないため、ユーザ類似度が算出されていない。

なお、本実験では、 $threshold^{user} = 0.85$ とする。したがって、図 36(c) より、被験者 A の例では、被験者 C および被験者 D が類似ユーザとして扱われる。

フェーズ 3：比較評価用の行動履歴データベースの特定

提案方式の評価を行うために、フェーズ 1 で作成した、各被験者の行動履歴データベースから、比較評価のための行動履歴データベースを特定する。

ここでは、対象とする被験者を被験者 A としたとき、以下の特定されたデータから行動履歴データベースを作成する。

- (I) Ω_I ：被験者 A の全データ
- (II) Ω_{II} ：被験者 A のデータのうち、推薦時コンテキストに関連するデータ
- (III) Ω_{III} ：被験者 $\{B, C, D, E, F, G, H\}$ の全データ
- (IV) Ω_{IV} ：被験者 $\{B, C, D, E, F, G, H\}$ のデータのうち、推薦時コンテキストに関連するデータ
- (V) Ω_V ：タイプ (a) の類似被験者 $\{C, D, E, G, H\}$ のデータのうち、推薦時コンテキストに関連するデータ
- (VI) Ω_{VI} ：タイプ (b) の類似被験者 $\{C, D, E, G\}$ のデータのうち、推薦時コンテキストに関連するデータ
- (VII) Ω_{VII} ：タイプ (c) の類似被験者 $\{C, D\}$ のデータのうち、推薦時コンテキストに関連するデータ

上記の各行動履歴データベースを、各被験者について作成する。

(a) コンテキストを考慮しないユーザ類似度

被験者	A	B	C	D	E	F	G	H
A	-	0.749	0.911	0.908	0.912	0.794	0.925	0.900
B	0.749	-	0.879	0.775	0.824	0.920	0.860	0.775
C	0.911	0.879	-	0.939	0.962	0.914	0.951	0.945
D	0.908	0.775	0.939	-	0.927	0.836	0.915	0.910
E	0.912	0.824	0.962	0.927	-	0.845	0.961	0.928
F	0.794	0.920	0.914	0.836	0.845	-	0.843	0.849
G	0.925	0.860	0.951	0.915	0.961	0.843	-	0.912
H	0.900	0.775	0.945	0.910	0.928	0.849	0.912	-

(b) 推薦時コンテキストのみを考慮したユーザ類似度

被験者	A	B	C	D	E	F	G	H
A	-	0.706	0.877	0.906	0.888	0.773	0.908	NaN
B	0.706	-	0.834	0.781	0.773	0.918	NaN	0.714
C	0.877	0.834	-	0.877	0.915	NaN	0.925	0.849
D	0.906	0.781	0.877	-	NaN	0.841	NaN	0.878
E	0.888	0.773	0.915	NaN	-	0.804	0.941	0.848
F	0.773	0.918	NaN	0.841	0.804	-	0.789	0.820
G	0.908	NaN	0.925	NaN	0.941	0.789	-	0.845
H	NaN	0.714	0.849	0.878	0.848	0.820	0.845	-

(c) 推薦時コンテキストと時系列行動履歴を考慮したユーザ類似度

被験者	A	B	C	D	E	F	G	H
A	-	0.626	0.870	0.882	0.838	0.702	0.825	NaN
B	0.626	-	NaN	0.715	NaN	0.902	NaN	0.704
C	0.870	NaN	-	NaN	0.941	NaN	0.890	0.823
D	0.882	0.715	NaN	-	NaN	0.825	NaN	0.857
E	0.838	NaN	0.941	NaN	-	0.769	0.898	0.783
F	0.702	0.902	NaN	0.825	0.769	-	NaN	0.803
G	0.825	NaN	0.890	NaN	0.898	NaN	-	NaN
H	NaN	0.704	0.823	0.857	0.783	0.803	NaN	-

図 36 本実験における被験者間のユーザ類似度

表 26 被験者 A の評価データに対する行動予測スコア (推薦時コンテキスト=「平日の 20 時に一人で」, 過去コンテキスト=「(19 時, ”食事する”)」の場合)

SID	日時	行動	評価	スコア						
				$Score_I^{SPM}$	$Score_{II}^{SPM}$	$Score_{III}^{SPM}$	$Score_{IV}^{SPM}$	$Score_V^{SPM}$	$Score_{VI}^{SPM}$	$Score_{VII}^{SPM}$
1	2008/12/1 20:00	食事する	1	0.000	0.000	0.022	0.000	0.017	0.017	0.017
2	2008/12/2 20:00	食事する	1	0.000	0.000	0.022	0.000	0.017	0.017	0.017
3	2008/12/3 20:00	食事する	1	0.000	0.000	0.022	0.000	0.017	0.017	0.017
4	2008/12/4 20:00	食事する	1	0.000	0.000	0.022	0.000	0.017	0.017	0.017
5	2008/12/5 20:00	お茶する	4	0.500	0.640	0.221	0.182	0.271	0.271	0.271
6	2008/12/6 20:00	お茶する	5	0.500	0.640	0.221	0.182	0.271	0.271	0.271
9	2008/12/9 20:00	バーで飲む	1	0.417	0.320	0.515	0.443	0.424	0.424	0.424
13	2008/12/13 20:00	買物する	2	0.135	0.240	0.159	0.205	0.305	0.305	0.305
17	2008/12/17 20:00	遊ぶ	1	0.167	0.000	0.162	0.114	0.000	0.000	0.000
20	2008/12/20 20:00	遊ぶ	1	0.167	0.000	0.162	0.114	0.000	0.000	0.000

フェーズ 4 : 比較評価用の行動履歴データベースに基づく行動予測スコアの算出
 フェーズ 3 で特定した各行動履歴データベース $\Omega_I, \Omega_{II}, \dots, \Omega_{VII}$ に基づいて、行動予測スコアを算出する。ここでは、フェーズ 1 において、各被験者がコンテキストごとに評価付けした施設データ (図 35(b)) (以降、評価データ) に対応した行動予測スコアを算出する。

行動予測スコアは、6.4 節で説明した、各方法により算出される。シーケンシャルパターンマイニング手法 (以降、SPM)(6.4.1 項) により、 Ω_i に基づいて、算出された行動予測スコアを $Score_i^{SPM}$ と表記する。同様に、LCS 抽出手法 (以降、LCS) (6.4.2 項) による行動予測スコアを $Score_i^{LCS}$ と表記する。

表 26 に、例として被験者 A の評価データのうち、推薦時コンテキストが「平日の 20 時に一人で」、過去コンテキストが「(19 時, ”食事する”)」であるときの、20 件の行動に対する SPM による予測スコアを示す。例えば、表 26 の $SID = 6$ の行動 “お茶する” に対しては、 $Score_I^{SPM} = 0.500, Score_{II}^{SPM} = 0.640, \dots, Score_{VII}^{SPM} = 0.271$ が算出されている。同様に、各被験者の各コンテキストにおけるスコアを算出する。

フェーズ 5 : スコア算出結果に対する評価

フェーズ 4 で算出された各スコアに基づいて、各被験者の各コンテキストにおけるデータをランキングする。コンテキストごとのデータ数は 20 個であるので、第 1 位から 20 位のランキングを得る。各スコアによるランキング結果を比較することにより、提案方式の有効性の評価および考察を行う。

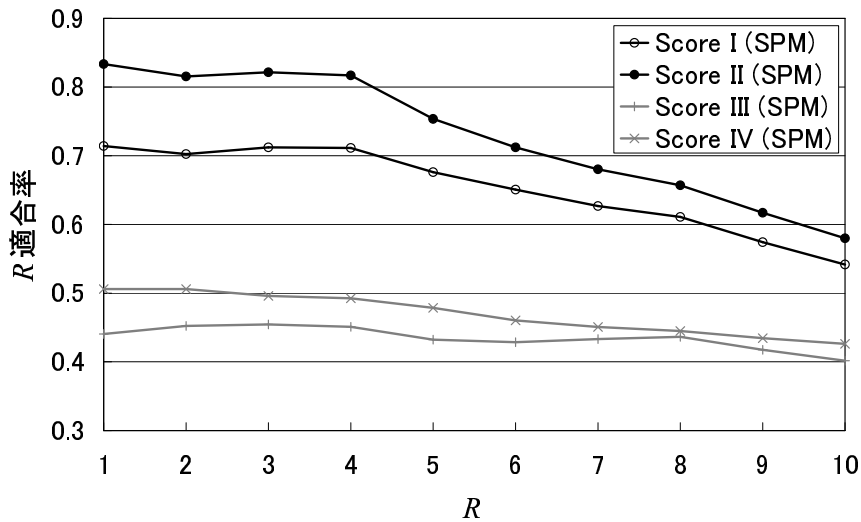


図 37 推薦時コンテキスト考慮時/非考慮時における SPM によるランキング結果に対する R 適合率

本実験では、評価尺度として、 R 適合率 [5] を用いる。

R 適合率は、上位から R 番目までのランキング結果を出力した時点での適合率のことである。なお、適合率は次式により表される。

$$\text{適合率} = \frac{\text{ランキング結果に含まれる適合データの数}}{\text{ランキング結果に含まれるデータの数}} \quad (64)$$

なお、被験者は各データに対し、5 段階の評価付けを行っているが、ここでは、5 段階評価のうち 4 以上のデータを適合データとして扱う。

6.5.3 実験結果および考察

ユーザの行動パターン抽出において推薦時コンテキストを考慮することの有効性に関する検証

図 37 は、SPM による行動予測スコア $Score_I^{SPM}$, $Score_{II}^{SPM}$, $Score_{III}^{SPM}$, $Score_{IV}^{SPM}$ に基づくランキング結果に対する第 10 位までの R 適合率である。ここでは、全被験者の全コンテキストにおけるデータを対象に求められた R 適合率を平均したものを示している。なお、SPM における最小サポート値は、 $\text{minsup} = 0.01$ と

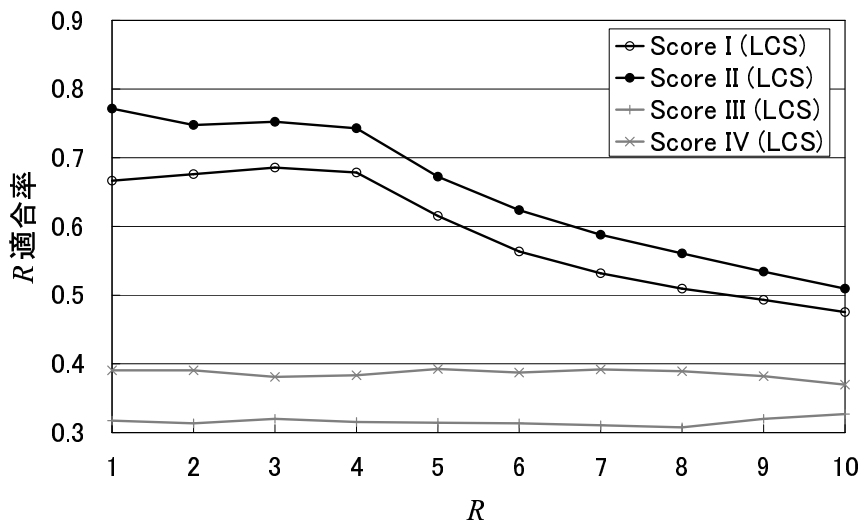


図 38 推薦時コンテキスト考慮時/非考慮時における LCS によるランキング結果に対する R 適合率

した。

$Score_I^{SPM}$, $Score_{II}^{SPM}$ は、対象被験者自身の行動履歴データベースに基づくスコアであり、 $Score_I^{SPM}$ では推薦時コンテキストを考慮せず、 $Score_{II}^{SPM}$ では考慮している。 $Score_{III}^{SPM}$, $Score_{IV}^{SPM}$ は、対象被験者以外の全被験者の行動履歴データベースに基づくスコアであり、 $Score_{III}^{SPM}$ では推薦時コンテキストを考慮せず、 $Score_{IV}^{SPM}$ では考慮している。

図 37 より、対象被験者自身の行動履歴データベースに基づくスコア、他被験者の行動履歴データベースに基づくスコア、いずれにおいても、推薦時コンテキストを考慮した方が、高い R 適合率を示した。この結果より、ユーザの行動履歴データベースから行動パターンを抽出する際に、推薦時コンテキストを考慮することの有効性を確認することができた。

同様に、LCS による結果を図 38 に示す。なお、LCS におけるパラメタは、 $min.Count = 1$, $min.Length = 2$ とした。図 38 より、LCS による結果についても SPM のときと同様のことを確認した。この結果より、提案方式が適用する行動パターン抽出手法に依存せず、有効に働くものであるということがいえる。

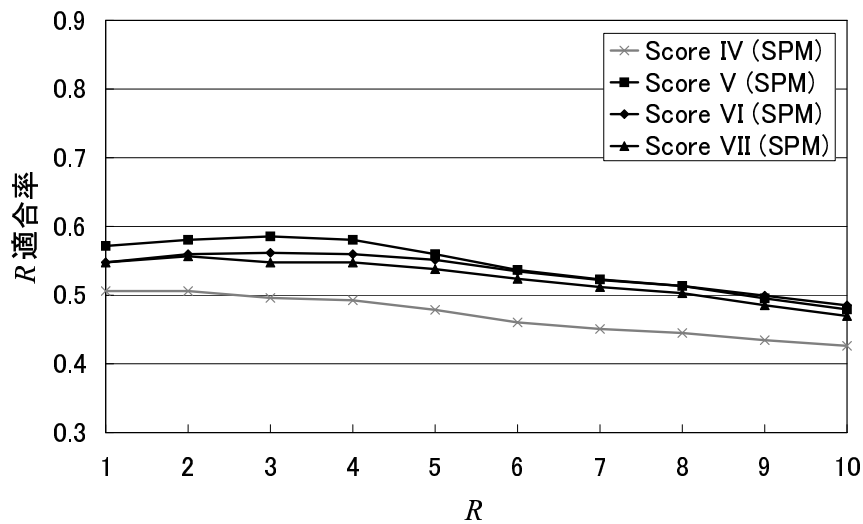


図 39 類似ユーザを考慮したときの SPM によるランキング結果に対する R 適合率

また、図 37 より、全体的に、他被験者の行動履歴データベースを参照するより、対象被験者自身の行動履歴データベースを参照した方が、 R 適合率が高くなった。これは、そもそもスコア算出に参照された行動履歴データベースが、被験者の評価用のデータに基づいて作成されているため、自明の結果であるといえる。したがって、対象ユーザ自身の行動履歴データが十分に蓄積されているという条件下においては、対象ユーザ自身の行動履歴データベースを参照した方が、より正確な推薦を行うことができるといえる。

ただし、実際には、対象ユーザ自身の行動履歴データが十分に確保できるとは限らない。このような場合、他ユーザの行動履歴データベースを参照することにより、データの不足を補うことができる。また、他ユーザの行動履歴データベースを参照することにより、自身のデータからだけでは気付かなかったような、新たな発見を促すことが期待できるということから、情報推薦において他ユーザの行動履歴データベースを参照することは有益であるといえる。

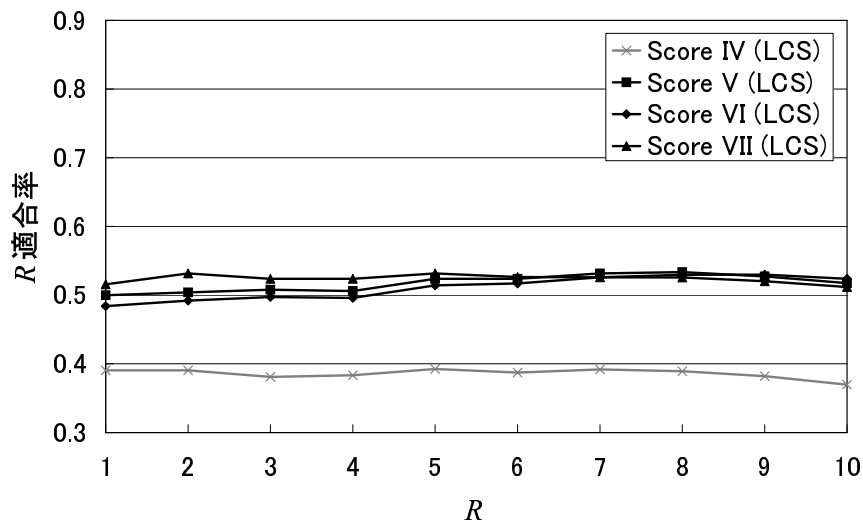


図 40 類似ユーザを考慮したときの LCS によるランキング結果に対する R 適合率

ユーザの行動パターン抽出において類似ユーザの行動履歴データを参照することの有効性に関する検証

図 39 は、SPM による行動予測スコア $Score_{IV}^{SPM}$, $Score_V^{SPM}$, $Score_{VI}^{SPM}$, $Score_{VII}^{SPM}$ に基づくランキング結果に対する第 10 位までの R 適合率である。

$Score_{IV}^{SPM}$ は、類似ユーザを考慮せず、他の全被験者の行動履歴データベースに基づくスコアである。一方、 $Score_V^{SPM}$, $Score_{VI}^{SPM}$, $Score_{VII}^{SPM}$ は、それぞれ、6.5.2 項「フェーズ 2」で述べた、各タイプのユーザ類似度により求められた類似被験者の行動履歴データベースに基づくスコアである。

同様に、LCS による結果を図 40 に示す。

図 39 および図 40 より、類似ユーザを考慮した方が、 R 適合率が高くなった。したがって、ただ他ユーザの行動履歴データベースを参照するだけでなく、嗜好が類似するユーザを対象に行動履歴データベースを参照することにより、推薦の正確性を向上させることができるといえる。

ただし、類似ユーザとして 3 タイプのものをを用いたが、本実験においては、各類似ユーザを用いたときの R 適合率には大きな違いは見られなかった。類似ユーザの求め方の違いによる有効性に関しては、今後、提案方式の実用化に向けた実

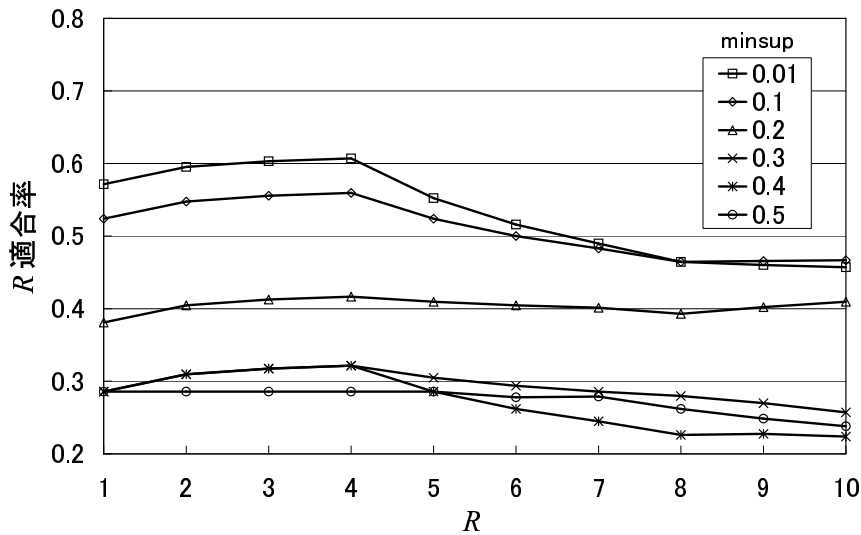


図 41 SPM における最小サポート値による R 適合率

証実験により検証していく必要がある。

SPM における最小サポート値の違いによる R 適合率の比較

図 41 は、SPM において最小サポート値を $\text{minsup} = \{0.01, 0.1, 0.2, 0.3, 0.4, 0.5\}$ に変えたときの、 R 適合率を比較したものである。算出スコアは $Score_V^{SPM}$ を対象としている。

図 41 のように、 $\text{minsup} = 0.01$ としたとき、 R 適合率が最も高く、 minsup を大きくするにつれて、 R 適合率が低くなった。これは、 minsup を小さくすることにより、多くの行動パターンを抽出することができるようになり、結果として対象被験者の行動と適合するものが多くなったと考えられる。したがって、本実験結果からでは、 minsup はできるだけ小さくすることが望ましいといえる。

ただし、6.4.1 項で述べたように、適切な minsup は、行動パターン抽出の処理時間や推薦結果に対する正確性や発見性など、何を重視するかによって変わってくる。これは、推薦システムを設計する際に決定すべきものであるため、今後、実証実験を通じて、適切な minsup について検討する。

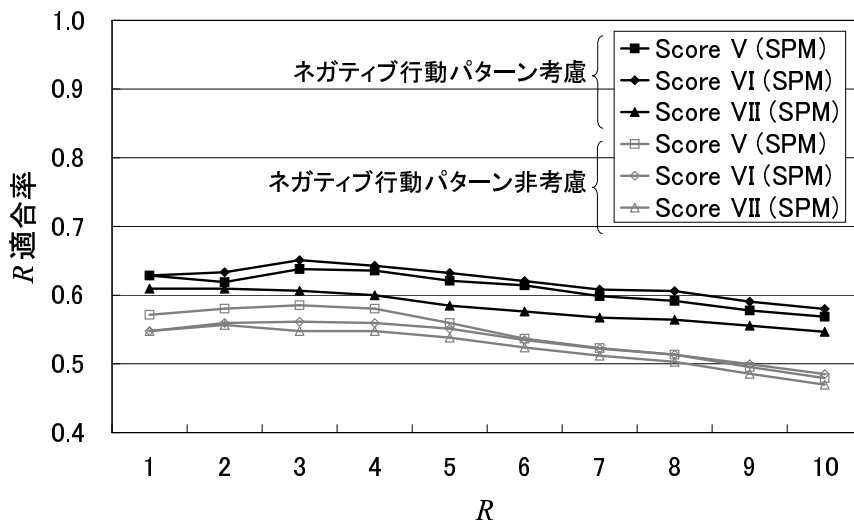


図 42 ネガティブ行動パターン考慮/非考慮時の R 適合率

行動予測スコア算出においてネガティブ行動パターンを考慮することの有効性に関する検証

本実験では、ユーザの行動履歴データベースを、図 35 のように、被験者が評価値 4 以上を与えた評価データに基づいて作成した。そして、この行動履歴データベースから行動パターンを抽出した。本項では、別途、被験者が評価値 2 以下を与えた評価データから行動履歴データベースを作成する。そして、同様にこの行動履歴データベースから行動パターンを抽出する。ここで抽出された行動パターンは、被験者にとって否定的なものであり、この否定的な行動パターンを考慮することによって、 R 適合率を向上させることができるか検証する。

ここでは、評価値 4 以上のデータに基づく行動履歴データベースから抽出された行動パターンをポジティブ行動パターン、評価値 2 以下のデータに基づく行動履歴データベースから抽出された行動パターンをネガティブ行動パターンとよび、各行動パターンに基づくスコアを算出する。ポジティブ行動パターンに基づくスコアを $Score^+$ 、ネガティブ行動パターンに基づくスコアを $Score^-$ としたとき、行動予測スコアを、

$$Score = Score^+ - Score^- \quad (65)$$

として求める。

図 42 は，SPM による行動予測スコア $Score_V^{SPM}$ ， $Score_{VI}^{SPM}$ ， $Score_{VII}^{SPM}$ に基づくランキング結果に対する第 10 位までの R 適合率である。図 42 のように，ネガティブ行動パターンを考慮したスコアと考慮していないスコアに基づいて R 適合率を求めている。

図 42 のように，いずれのスコアにおいても，ネガティブ行動パターンを考慮したときの方が R 適合率が高くなった。この結果より，このネガティブ行動パターンに基づいて，行動予測スコアを算出する。そして，行動予測スコア算出においてネガティブ行動パターンを考慮することにより，推薦の性能を向上させることができるといえる。

6.6 結言

本章では，ユーザの推薦時点の状況に加え，ユーザがその時点までに取ってきた行動および，その時点以降に予定している行動もユーザコンテキストとして扱った，コンテキスト依存型情報推薦方式を提案した。提案方式では，ユーザの行動履歴データベースから，推薦時コンテキストに応じた限定的な行動パターンの抽出を行い，この限定的な行動パターンに基づき，推薦情報の判定を行う。推薦時コンテキストに応じた限定的な行動パターンを考慮することにより，推薦時コンテキストおよびユーザの時系列的な行動履歴の両方を考慮した情報推薦を行うことを可能とした。

本章で得られた成果と知見を以下に示す。

- ユーザの行動パターン抽出において，推薦時コンテキストを考慮することの有効性を示した。
- ユーザの行動パターン抽出において，類似ユーザの行動履歴データを参照することの有効性を示した。
- シーケンシャルパターンマイニング手法を適用する際に，最小サポート値を小さくすることにより，よりユーザに適切な行動パターンを抽出することができることを示した。

- 行動予測スコア算出において，ユーザにとってネガティブな行動パターンを考慮することの有効性を示した．

今後は提案方式の実用性を検証するために，提案方式を，観光地における観光施設情報推薦システムや都市部における回遊情報推薦システムなどに適用した実証実験を行う．実証実験を通じて，ユーザ類似度の適切な算出方法について検討する．また，SPMにおける最小サポート値などのパラメタ調整についても検討する．

7. 結論

7.1 論文のまとめ

本論文では、ユーザの嗜好だけでなくそのときのユーザコンテキストに合った情報を提供する、コンテキスト依存型情報推薦システムを提案した。コンテキスト依存型情報推薦システムを実現するための基礎的技術として、

- (1) コンテキスト依存型ユーザ嗜好モデルに基づく情報推薦方式
- (2) コンテキスト依存型ランキング方式
- (3) ユーザコンテキストの時系列性を考慮した情報推薦方式

を提案し、その有効性の検証を行った。

本論文第2章では、情報推薦技術およびコンテキストウェアシステムの従来に関連研究を整理し、本研究の位置付けについて述べた。

本論文第3章では、本論文におけるユーザコンテキストを「情報選択過程におけるユーザの嗜好に影響を及ぼす、ユーザのおかれている状況」と定義した。また、本論文で提案するコンテキスト依存型情報推薦システムの全体像を示し、上記(1)~(3)のコンテキスト依存型情報推薦方式を実現するための各基礎的技術の位置付けを明確にした。

本論文第4章では、コンテキスト依存型ユーザ嗜好モデルおよびこのモデルに基づいた情報推薦方式を提案した。また、提案方式を飲食店情報推薦に対して適用した検証実験を行った。この検証実験より得られた成果と知見は以下のとおりである。

- 他のクラス分類手法との比較評価の結果、SVMは未学習データに対する汎化性能に優れていることが実証でき、情報推薦方式に対して適用することの妥当性を裏付けた。
- 二つの情報推薦方式、C-IFおよびC-CFに対して以下の特長を明らかにした。
 - 既知エリアを推薦対象とした場合には、C-IFによる情報推薦が有効に働く。

- 未知エリアを推薦対象とした場合には，C-CF による情報推薦が有効に働く．

- 推薦対象のアイテムデータおよびユーザコンテキストの特徴パラメタの最適化に関する検討を行い，各特徴パラメタが識別性能に与える影響を分析することによって，特徴パラメタの最適化が可能であるということを示した．

本論文第 5 章では，コンテキスト依存型価値判断基準モデルおよびこのモデルに基づいたランキング方式を提案した．また，被験者実験により，提案方式の有効性を検証した．本実験により得られた成果と知見は以下のとおりである．

- 提案方式では，ユーザコンテキストに応じて変化するユーザの価値判断基準を適切にモデル化することにより，ユーザの直面しているコンテキストにおいて相応しいランキング結果を提供することができる．
- ユーザのアイテムパラメタに対する重要度に基づいて，ユーザが重要視しているアイテムパラメタを適切に選定することにより，ランキング結果に対するユーザ満足度をより向上させることができる．
- コンテキストパラメタのユーザ嗜好に対する影響度に基づいて，影響の大きいコンテキストパラメタのみに着目し学習データを特定することにより，ユーザの直面しているコンテキストにおいて，より相応しいランキング結果を提供することができる．

本論文第 6 章では，コンテキスト依存型行動モデルをおよびこのモデルに基づいた情報推薦方式を提案した．そして，被験者実験を通じて，提案方式の有効性について検証および考察を行った．本実験により得られた成果と知見は以下のとおりである．

- 推薦時コンテキストに限定した行動パターンを抽出することにより，ユーザにとって適切な情報推薦が可能であることを示した．
- 類似ユーザ選定の際に，ユーザ類似度の正確性と確保できる類似ユーザ数を考慮することが重要であることを確認した．

7.2 コンテキスト依存型情報推薦システムの適用例

本研究で提案したコンテキスト依存型情報推薦システムは、飲食店や音楽、観光情報など、ユーザが目的の情報を選択する際に、ユーザのコンテキスト依存の嗜好があらわれるようなコンテンツに適用可能なものである。以下、本推薦システムをいくつかのコンテンツに適用したときの例を示す。

飲食店情報推薦システム

ある日、Aさんが朝早くに一人で街を歩いていた。このとき、Aさんは、携帯電話を用い、推薦システムに対し、周辺にある飲食店の情報を要求した。すると、推薦システムは、Aさんの嗜好と、そのときのコンテキストを考慮し、落ち着いた雰囲気のカフェをAさんに推薦した。

また、別の日の夜に、Aさんは恋人と繁華街で食事を済ませていた。このとき、また次の店を探すため、推薦システムに周辺の飲食店の情報を要求した。すると、推薦システムは、Aさんのこれまでの経験を考慮し、夜景の見えるバーをAさんに推薦した。

従来の飲食店検索サービスを用いた方法では、ユーザが飲食店に対する条件を細かく指定する必要があった。これに対し、本推薦システムでは、ユーザのこれまでの経験から、そのコンテキストにおけるユーザの嗜好を自動的に推測するため、ユーザからの細かい入力を必要としない。

音楽推薦システム

Bさんは、携帯型音楽プレイヤーを用い、毎日、さまざまな場面で音楽を聴いている。一人でドライブをしているときには、テンポの速いポップスが自動的に流れる。恋人とドライブをしているときには、バラード系の音楽が再生される。また、Bさんが部屋で勉強しているときには、集中しやすいように、モーツァルトの音楽が流れる。

近年の携帯型音楽プレイヤーは、大容量化が進み、膨大な楽曲データを持ち運ぶことができるようになった。その一方で、その膨大な楽曲データの中から、聴きたい曲を選ぶには、ユーザにとって大変手間となる。こうした手間から、再生す

る曲もマンネリに陥る可能性もある。これに対し、本推薦システムを用いることで、ユーザの楽曲選択における手間を省くことができ、さらに、ユーザが聴いたことのないような楽曲も再生される可能性もあり、ユーザにとって新たな興味の発見にもつながる。

観光情報推薦システム

ある日、Cさんは、子供を連れて、奈良の東大寺を訪れた。次に、どこに行くか決めるため、推薦システムに奈良の観光情報を要求した。すると、推薦システムは、Cさんに対し、子供が鹿と戯れることができる、若草山を推薦した。

また別のDさんは、夕方に恋人とデートで東大寺を訪れた。Dさんは、夜に訪れる場所を決めるため、推薦システムに観光情報を要求した。すると、推薦システムは、デート向けに、夜のライトアップが綺麗な猿沢池を推薦した。

観光情報に関しては、インターネットなどで入手できるようになったが、いつ誰と観光するかによって、どの観光施設を訪問すれば良いかは変わってくる。本推薦システムでは、ユーザのコンテキストに合った観光施設を推薦することができる。

7.3 コンテキスト依存型情報推薦システムの運用指針

本節では、このコンテキスト依存型情報推薦システムの運用指針について述べる。

まず、推薦システム設計者（以下、設計者）は、システム運用開始前に、以下の手順を踏む。

- (1) 推薦対象アイテムおよびユーザコンテキストに関する初期の特徴パラメタを列挙する。
- (2) 列挙した特徴パラメタのタイプに応じてパラメタ表現方式を決める。
- (3) モニタを対象にした試験運用による特徴パラメタの最適化処理を行う。

まず，適用するコンテンツを決定したとき，設計者が下記の留意点を踏まえた上で，推薦対象アイテムおよびコンテキストの特徴パラメタを列挙する（例として，第4章表4および表5参照）．

- アイテム選択を行う上で，ユーザの嗜好に影響を及ぼすものであるか．
- どのようなユーザ層（ビジネスマンや主婦など），どのような場面（日常生活や旅行先など）をターゲットとするのか．

なお，列挙する特徴パラメタの個数については特に条件はない．推薦対象アイテムやコンテキストを十分に表現できるのであれば，最小は1個のパラメタでも良い．一方で，特徴パラメタは後に最適化処理を行うため，列挙すべきパラメタの個数に上限はない．

つづいて，本研究で扱う特徴パラメタは，そのタイプに応じて次のパラメタ型により表現される（第5章図14参照）．

- 連続値型パラメタ
- 周期型パラメタ
- 2値型パラメタ
- カテゴリ型パラメタ

設計者は，列挙した特徴パラメタを，各パラメタのタイプに応じて上記のパラメタ型により表現する．

必要と思われる特徴パラメタを列挙し，各パラメタ型への変換が終われば，列挙した全パラメタを用いて，試験用の推薦システムを実装する．そして，一般公募したモニタを対象とした試験運用により特徴パラメタの最適化処理を行う（4.7節参照）．手順は以下のとおりである．

- (1) 一定期間，モニタに推薦システムを利用してもらいながら，モニタのユーザ嗜好モデルを構築する．
- (2) モニタのモデルを用い，特徴パラメタ最適化のための実験を行い，最終的に実システムで利用するパラメタを決定する．

(3) 最適化済みの特徴パラメタを用い、モニタのユーザ嗜好モデルを再構築する。

ここで、モニタの人数や試験運用期間などについては、今後の実用化に向けた実証実験を通じて決定するものとする。

最終的に、最適化済みの特徴パラメタを用い、推薦システムを実装し、一般ユーザへの運用を開始する。このとき、初期のユーザは自身の学習データ数が乏しいため、他ユーザのモデルを参照する C-CF による推薦を優先的に行う(4.6.3 節参照)。特に、システム運用初期には、全ユーザの学習データ数が乏しいため、モニタのモデルを参照した C-CF による推薦を行う。対象ユーザの学習データ数が十分に増えてくれば、対象ユーザのモデルに基づいた C-IF による推薦も行う。ここで、学習データ数に応じた C-IF/C-CF による推薦の優先度の比率は、実証実験を通じて調整するものとする。さらに、推薦時コンテキストや過去/未来コンテキストを考慮したランキングを提供することにより、よりユーザ満足度の高い推薦を行うことを可能とする(第5章および第6章参照)。

7.4 ユーザのデモグラフィック情報の利用

デモグラフィックとは、人口統計学的な属性データであり、具体的には、性別、年齢、居住地域、所得、職業、学歴、家族構成などである。マーケティングにおいては、これらの属性を分析し、消費者を複数のコミュニティに分類した上で、当該コミュニティに適切なサービスを提供するといった方法が取られている。例えば、20代男性の人に商品を推薦する場合、一般に20代男性が好んでいる商品を提供する。

しかし、人々の生活スタイルが多様化するにつれ、人の嗜好も多様になってきた。このためデモグラフィック情報が一致するからといって、必ずしも嗜好も一致するとはいえなくなってきた。個々人の多様な嗜好に対応するためには、一般的な嗜好にとらわれず、個人ごとの嗜好を適切にモデル化する必要があると考えられる。

本研究では、この考えに基づき、デモグラフィック情報は用いず、おかれているコンテキストの情報およびアイテムの情報のみを用いて、ユーザの嗜好を表現

している．こうすることにより，個人に特化した嗜好モデルを構築することを可能としている．

7.5 今後の展望

本研究では，提案の推薦方式に関する基礎評価を行い，その有用性を示した．今後の展望としては，第4章～第6章で提案した各基礎技術を統合した，コンテキスト依存型情報推薦システムを構築する．そして，飲食店や音楽情報，観光情報など，さまざまなコンテンツを対象とした実証実験を行い，コンテキスト依存型情報推薦システムの実用化に向けた検討を行う．

ユーザが取得できる情報量が膨大化，さらにはユーザ嗜好そのものが多様化している中，コンテキスト依存型情報推薦システムを確立したことに対する意義は大きい．

ユビキタスコンピューティング環境の整備に伴い「いつでも」「どこでも」「だれでも」情報技術の恩恵を受けることができるようになる．このような中，個々のユーザの嗜好，その時，その場のユーザコンテキストに合った情報を提供する，コンテキスト依存型情報推薦システムにより，人々の生活をより安全・安心・快適なものにすることができるようになることを期待できる．

さらには，今後は情報技術がより自然な形でユーザを支援できるように，人間と調和する情報環境の構築が求められるようになると思われる．人間にとってより良いサービスを提供するためには，人間の行動や嗜好を適切にモデル化することが重要となってくる．本論文では，ユーザモデルの一つとして，ユーザコンテキストに応じて多種多様に変化するユーザ嗜好を適切にモデル化する技術を確立した．この技術により，ユーザの嗜好およびユーザコンテキストに応じたサービスが提供されるようになり，人々の生活をより豊かなものにできると期待できる．

謝辞

本研究の遂行並びに論文の執筆にあたり，主指導教員であるインタラクティブメディア設計学講座兼データベース学講座 加藤 博一 教授には，懇切なる御指導，御鞭撻を賜りました．また，日々研究に打ち込めるよう快適な研究環境を御提供下さいました．謹んで感謝の意を表します．

奈良産業大学 植村 俊亮 教授には，本研究の遂行にあたり，貴重な御助言を頂きました．また，私が博士前期課程2年次の途中から，データベース学講座に転籍した際には，私を温かく迎え入れて下さり，退官された後も，私を温かく見守って下さいました．ここに深く御礼申し上げます．

知能情報処理学講座 木戸出 正繼 教授には，本研究の副指導教員になって頂き，本論文の作成にあたり，貴重な御助言を賜りました．また，私が知能情報処理学講座に在籍していたときから，研究への取り組み方，社会人としての心構えについて厳しくも温かい御指導を頂き，私の成長を見守って下さいました．ここに厚く御礼申し上げます．

インタラクティブメディア設計学講座兼データベース学講座 宮崎 純 准教授には，本研究の副指導教員になって頂き，本研究の遂行並びに本論文の執筆にあたり，懇切なる御指導，御鞭撻を賜りました．特にデータベース学の専門的な観点から，有益な御助言を頂きました．ここに深く御礼申し上げます．

京都産業大学 中島 伸介 准教授には，本研究の直接指導教員として，本研究の遂行並びに論文の執筆にあたり，厳しくも温かい御指導，御鞭撻を賜りました．また，日頃より研究者として社会人として，そして人間としての心構えについても，多大な御指導を頂きました．何よりも，研究者としての魂の重要性について熱心に御教示頂きました．ここに深く感謝の意を表し，厚く御礼申し上げます．

大阪大学 土方 嘉徳 先生には，学外審査委員として，有益な御助言と御教示を頂きました．また，学外の学会及び研究会等におきまして，本研究を遂行するにあたり，情報推薦技術の専門的な観点から，多くの貴重な御助言を賜りました．謹んで感謝の意を表します．

本研究の遂行にあたり，学外では，学会や研究会等におきまして，多くの先生方に御助言を賜りました．京都大学 田中 克己 教授，兵庫県立大学 角谷 和俊 教

授には、貴重な御助言と御教示を賜ると共に、激励し続けて頂きました。ここに深く感謝の意を表し、厚く御礼申し上げます。

学内では、ソフトウェア基礎学講座 伊藤 実 教授，安本 慶一 准教授には、貴重な御教示を賜ると共に、激励を頂きました。謹んで感謝の意を表します。

知能情報処理学講座 浮田 宗伯 准教授には、私が博士前期課程2年次の途中に、知能情報処理学講座からデータベース学講座に転籍する際に、研究並びに進路についての悩みに対して、親身になって御相談に乗って頂きました。心より感謝の意を表し、厚く御礼申し上げます。

所属講座であるインタラクティブメディア設計学講座兼データベース学講座では、波多野 賢治 先生 (現、同志社大学)，天野 敏之 先生，藤澤 誠 先生には、日頃より、講座内研究会におきまして、有益な御助言と御教示を頂きました。厚く御礼申し上げます。様々な面で研究活動を支えて下さった、秘書 上野 真紀子 氏，元秘書 今田 薫 氏に、謹んで感謝の意を表します。同輩，栗田 裕人 氏，浜野 泰男 氏をはじめ、諸先輩，後輩の皆様には、本研究において多大な御協力を頂きました。また、公私にわたって共に励まし合うことができました。心より感謝致します。

私がデータベース学講座に転籍する際に、研究協力部 学生課 刀谷 敏博 氏には、進路についての悩みに対して、親身になって御相談に乗って頂きました。ここに深く御礼申し上げます。

知能情報処理学講座在籍時を含め、大学院生活において激励し続けて頂きました。知能情報処理学講座の諸先輩，同輩，後輩及び秘書の皆様には心より感謝致します。特に、研究の悩みに対して親身になって御相談に乗って頂いた、満上 育久 氏 (現、京都大学) には厚く御礼申し上げます。

大阪市立大学 土木計画学研究室 日野 泰雄 教授，内田 敬 准教授，吉田 長裕 先生には、研究者としての原点として、研究への取り組み方，研究者としての心構えを御教示頂きました。心より感謝の意を表します。特に、内田准教授には、私の研究に対する興味を引き出して頂きました。また、早期から多くの論文発表の機会を与えて頂き、研究を進める面白さ，困難さを味わうとともに、それを乗り越えた際の喜びを経験することができました。ここに厚く御礼申し上げます。土

木計画学分野において、苦楽を共に過ごし、お互い励まし合ってきた、諸先輩、同輩、後輩、秘書の皆様に心より感謝致します。

同期生として共に励まし合ってきた、応用システム科学講座 中村 幸紀 氏に心より感謝の意を表します。論理生命学講座 井原 瑞希 氏には、本研究に興味を持って頂き、共に学内研究プロジェクト CICP (Creative and International Competitiveness Project) を遂行する機会を提供頂きました。ここに感謝の意を表します。

大阪市立大学ソフトテニス部及び奈良先端科学技術大学院大学 テニスサークル FESTY では、様々な考え方を持った多くの方々と出会うことができ、研究の合間の憩いの場として、皆様と楽しいひとときを過ごすことができました。また、チームプレーの大切さを学ぶことができました。メンバの皆様に心より感謝致します。

最後に、父 奥 青児、母 奥 房子、姉 奥 静愛は、私を温かく見守り、支え続けてくれました。愛犬 ナナ、リリ、ロロは、私に癒しを与えてくれました。家族に心より感謝します。

平成 21 年 3 月 奥 健太

参考文献

- [1] 『amazon.com』 . <http://www.amazon.com/> (2008-12-14 確認).
- [2] 『google news』 . <http://news.google.com/> (2008-12-14 確認).
- [3] 土方嘉徳. 情報推薦・情報フィルタリングのためのユーザプロファイリング技術. 人工知能学会誌, Vol. 19, No. 3, pp. 365–372, 2004.
- [4] 土方嘉徳. 1. 嗜好抽出と情報推薦技術 (嗜好抽出・情報推薦の基礎理論, 特集; 利用者の好みをとらえ活かす-嗜好抽出技術の最前線-). 情報処理学会誌, Vol. 48, No. 9, pp. 957–965, 9 2007.
- [5] 北 研二 (著), 津田 和彦 (著), 獅々堀 正幹 (著). 『情報検索アルゴリズム』 . 共立出版, 2001.
- [6] J. J. Rocchio. Relevance feedback in information retrieval. In *The SMART Retrieval System: Experiments in Automatic Document Processing*, pp. 313–323. Prentice Hall, 1971. Salton, G. (Ed.).
- [7] G. Salton, A. Wong, and C. S. Yang. A vector space model for automatic indexing. *Communications of the ACM*, Vol. 18, No. 11, pp. 613–620, 1975.
- [8] J. Weston and C. Watkins. Multi-class support vector machines. Technical report, Technical report csd-tr-98-04, Royal Holloway, University of London, Surrey, England, 1998.
- [9] J. Weston and C. Watkins. Support vector machines for multi-class pattern recognition. In *M. Verleysen (ed.) Proceedings of the 7th European Symposium on Artificial Neural Networks (ESANN-99)*, pp. 219–224, Bruges, Belgium, 1999.
- [10] 篠田裕之, 竹内亨, 寺西裕一, 春本要, 下條真司. ユビキタス環境における協調フィルタリングを用いた行動ナビゲーション手法の考察. 情報処理学会研

- 究報告, マルチメディア通信と分散処理研究会報告, IPSJ SIG Notes 2007-DPS-130-(14)), 第 2007 巻, pp. 77–82, 9 2007.
- [11] 山元理絵, 小林大, 吉原朋宏, 小林隆志, 横田治夫. アクセスログに基づく web ページ推薦における lcs の利用とその解析. 情報処理学会論文誌: データベース (TOD34), Vol. 48, No. SIG 11, pp. 38–48, jun 2007.
- [12] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. GroupLens: An open architecture for collaborative filtering of netnews. In *Proc. of ACM 1994 Conference on Computer Supported Cooperative Work (CSCW'94)*, pp. 175–186, 1994.
- [13] M. Balabanovic and Y. Shoham. Fab: Content-based, collaborative recommendation. *Comm. ACM*, Vol. 40, No. 3, pp. 66–72, 1997.
- [14] Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. In *IEEE Trans. on Knowl. and Data Eng.*, Vol. 17, pp. 734–749. IEEE Educational Activities Department, 2005.
- [15] K. Lang. Newsweeder: Learning to filter netnews. Proc. 12th Int'l Conf. Machine Learning, 1995.
- [16] R. J. Mooney and L. Roy. Content-based book recommending using learning for text categorization. In *Proc. ACM SIGIR '99 Workshop Recommender Systems: Algorithms and Evaluation*, 1999.
- [17] M. Pazzani and D. Billsus. Learning and revising user profiles: The identification of interesting web sites. In *Machine Learning*, Vol. 27, pp. 313–331, 1997.
- [18] C. Cortes and V. Vapnik. Support-vector networks,. In *Machine Learning*, Vol. 20, pp. 273–297, 1995.

- [19] TM. Mitchell. *Machine Learning*. McGraw Hill, 1997.
- [20] M. Sahami, S. Dumaisy, D. Heckermany, and E. Horvitz. A bayesian approach to filtering junk e-mail. In *AAAI-98 Workshop on Learning for Text Categorization*, 1998.
- [21] B. S. Everitt. *Cluster Analysis*. Arnold, London, 1993.
- [22] J. R. Quinlan. Induction of decision trees. In *Machine Learning*, Vol. 1, pp. 81–106, mar 1986.
- [23] J. R. Quinlan. *The Morgan Kaufmann Series In Machine Learning*, chapter C4.5: Programs for Machine Learning. may 1993.
- [24] Simon Haykin. *NEURAL NETWORKS: a comprehensive foundation*. Prentice-Hall Engineering/Science/Mathematics, international editions edition, 1998.
- [25] M. Pazzani. A framework for collaborative, content-based, and demographic filtering. *Artificial Intelligence Rev*, pp. 393–408, dec 1999.
- [26] N. Littlestone and M. Warmuth. The weighted majority algorithm. *Information and Computation*, Vol. 108, No. 2, pp. 212–261, 1994.
- [27] 岩濱数宏, 土方嘉徳, 西田正吾. 決定木を用いた音楽情報フィルタリングシステムとその有効性の検証. 電子情報通信学会論文誌 D, Vol. J88-D1, No. 3, pp. 642–656, 2005.
- [28] J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Riedl. Grouplens: Applying collaborative filtering to usenet news. *Comm. ACM*, Vol. 40, No. 3, pp. 77–87, 1997.
- [29] W. Hill, L. Stead, M. Rosenstein, and G. Furnas. Recommending and evaluating choices in a virtual community of use. In *Proc. Conf. Human Factors in Computing Systems*, 1995.

- [30] U. Shardanand and P. Maes. Social information filtering: Algorithms for automating 'word of mouth'. In *Proc. Conf. Human Factors in Computing Systems*, 1995.
- [31] G. Linden, B. Smith, and J. York. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, jan / feb 2003.
- [32] 『tsutaya on line』 . <http://www.tsutaya.co.jp/> (2008-12-14 確認).
- [33] J.S. Breese, D. Heckerman, and C. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In *Proc. 14th Conf. Uncertainty in Artificial Intelligence*, jul 1998.
- [34] J. Delgado and N. Ishii. Memory-based weighted-majority prediction for recommender systems. In *Proc. ACM SIGIR '99 Workshop Recommender Systems: Algorithms and Evaluation*, 1999.
- [35] A. Nakamura and N. Abe. Collaborative filtering using weighted majority prediction algorithms. In *Proc. 15th Int'l Conf. Machine Learning*, 1998.
- [36] D. Billsus and M. Pazzani. Learning collaborative information filters. In *Proc. Int'l Conf. Machine Learning*, 1998.
- [37] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins. Eigentaste: A constant time collaborative filtering algorithm. *Information Retrieval J.*, Vol. 4, No. 2, pp. 133–151, jul 2001.
- [38] T. Hofmann. Collaborative filtering via gaussian probabilistic latent semantic analysis. In *Proc. 26th Ann. Int'l ACM SIGIR Conf.*, 2003.
- [39] B. Marlin. Modeling user rating profiles for collaborative filtering. In *Proc. 17th Ann. Conf. Neural Information Processing Systems (NIPS '03)*, 2003.

- [40] D. Pavlov and D. Pennock. A maximum entropy approach to collaborative filtering in dynamic, sparse, high-dimensional domains. In *Proc. 16th Ann. Conf. Neural Information Processing Systems (NIPS '02)*, 2002.
- [41] L. H. Ungar and D.P. Foster. Clustering methods for collaborative filtering. In *Proc. Recommender Systems, Papers from 1998 Workshop, Technical Report WS-98-08*, 1998.
- [42] M. Baldauf, , and S. Dustdar. A survey on context-aware systems. Technical report, Technical Report TUV-1841-2004-24, Technical University of Vienna, 2004.
- [43] R. Want, A. Hopper, V. Falcao, and J. Gibbons. The active badge location system. In *ACM Transactions on Information Systems*, Vol. 10, pp. 91–102. 1992.
- [44] Gregory D. Abowd, Christopher G. Atkeson, Jason Hong, Sue Long, Rob Kooper, and Mike Pinkerton. Cyberguide: A mobile context-aware tour guide. *Wireless Networks*, Vol. 3, No. 5, pp. 421–433, oct 1997.
- [45] Y. Sumi, T. Etani, S. Fels, N Simonet, K. Kobayashi, and K Mase. *Community Computing and Support Systems, Social Interaction in Networked Communities*, chapter C-map: Building a context-aware mobile assistant for exhibition tours, pp. 137–154. Springer-Verlag, London, UK, 1998.
- [46] K. Cheverst, N. Davies, K. Mitchell, and A. Friday. Experiences of developing and deploying a context-aware tourist guide: the guide project. In *Proceedings of the 6th Annual international Conference on Mobile Computing and Networking (MobiCom2000)*, pp. 20–31, ACM, New York, NY, 2000.
- [47] 森下健, 中尾恵, 垂水浩幸, 上林弥彦. 時空間限定オブジェクトシステム spacetag : プロトタイプシステムの設計と実装. *情報処理学会論文誌*, Vol. 41, No. 10, pp. 2689–2697, 2000.

- [48] Mauro Brunato, Roberto Battiti, Alessandro Villani, and Andrea Delai. A location-dependent recommender system for the web. Technical Report Technical Report DIT-02-093, Informatica e Telecomunicazioni, University of Trento, 2002.
- [49] Soe-Tsyr Yuan and Y. W. Tsao. A recommendation mechanism for contextualized mobile advertising. *Expert Systems with Applications*, Vol. 24, No. 4, pp. 399–414, May 2003.
- [50] Mark van Setten, Stanislav Pokraev, and Johan Koolwaaij. *Context-Aware Recommendations in the Mobile Tourist Application COMPASS*, Vol. 3137/2004 of *Lecture Notes in Computer Science*, chapter Adaptive Hypermedia and Adaptive Web-Based Systems, pp. 235–244. Springer Berlin / Heidelberg, 2004.
- [51] Oh Byung Kwon. “i know what you need to buy”: context-aware multimedia-based recommendation system. *Expert Systems with Applications*, Vol. 25, No. 3, pp. 387–400, October 2003.
- [52] Zhiwen Yu, Xingshe Zhou, Daqing Zhang, Chung-Yau Chin, Xiaohang Wang, and Ji Men. Supporting context-aware media recommendations for smart phones. *IEEE Pervasive Computing*, Vol. 5, No. 3, pp. 68–75, 2006.
- [53] Bill Schilit, Norman Adams, and Roy Want. Context-aware computing applications. In *Proceedings of IEEE Workshop on Mobile Computing Systems and Applications*, pp. 85–90, Santa Cruz, California, dec 1994. IEEE Computer Society Press.
- [54] N. Ryan, J. Pascoe, and D. Morse. Enhanced reality fieldwork: the context-aware archaeological assistant’. In *Proceeding of the 25th Anniversary Computer Applications in Archaeology*, 1997.
- [55] Anind K. Dey and Gregory D. Abowd. Cybrereminder: A context-aware system for supporting reminders. In *Proceedings of Second International Sym-*

- posium on Handheld and Ubiquitous Computing, HUC 2000*, pp. 172–186, Bristol, UK, sep 2000. Springer Verlag.
- [56] T. Strang and C. Linnhoff-Popien. A context modeling survey. In *First International Workshop on Advanced Context Modelling, Reasoning and Management, UbiComp*, 2004.
- [57] Guanling Chen and David Kotz. A survey of context-aware mobile computing research. Paper TR2000-381, Department of Computer Science, Dartmouth College, nov 2000.
- [58] A. Schmidt, M. Beigl, and H. W. Gellersen. There is more to context than location. *Comp. and Graphics J.*, Vol. 23, No. 6, pp. 893–902, dec 1999.
- [59] Albrecht Schmidt, Kofi Asante Aidoo, Antti Takaluoma, Urpo Tuomela, Kristof Van Laerhoven, and Walter Van de Velde. Advanced interaction in context. In *Proceedings of First International Symposium on Handheld and Ubiquitous Computing HUC'99*, pp. 89–101, Karlsruhe, Germany, sep 1999. Springer Verlag.
- [60] Anind K. Dey, Masayasu Futakawa, Daniel Salber, and Gregory D. Abowd. The conference assistant: Combining context-awareness with wearable computing. In *Proceedings of the 3rd International Symposium on Wearable Computers (ISWC '99)*, pp. 21–28, San Francisco, CA, oct 1999. IEEE Computer Society Press.
- [61] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification and Scene Analysis*. John Wiley & Sons, New York, 2nd ed. edition, 1993.
- [62] David J. Hand, Heikki Mannila, and Padhraic Smyth. *Principles of Data Mining*. The MIT Press, 2001.
- [63] C. C. Chang and C. J. Lin. *LIBSVM: a Library for Support Vector Machines*, 2006.

- [64] 『libsvm』 . <http://www.csie.ntu.edu.tw/~cjlin/libsvm/> (2008-12-14 確認).
- [65] 『weka』 . <http://www.cs.waikato.ac.nz/ml/weka/> (2008-12-14 確認).
- [66] 前田英作. 痛快!サポートベクトルマシン : 古くて新しいパターン認識手法. 情報処理, Vol. 42, No. 7, pp. 676–683, 2000.
- [67] 平博順. Svm のテキスト分類への応用. 2004 年電子情報通信学会総合大会, 2004.
- [68] 『yahoo! グルメ』 . <http://gourmet.yahoo.co.jp/> (2008-12-14 確認).
- [69] 加藤由花, 川口賢二, 箱崎勝也. オンラインショッピングを対象とした正確性と意外性のバランスを考慮したリコメンダシステム. 情報処理学会論文誌 : データベース (TOD27), Vol. 46, No. 13, pp. 53–64, 2005.
- [70] 中田豊久, 金井秀明, 國藤進. 実世界での利用を考慮した図書推薦モデルの提案と評価. 情報処理学会論文誌, Vol. 48, No. 1, pp. 148–162, 2007.
- [71] 李知雄, 西井和夫, 近藤勝直, 佐々木邦明. 都市観光回遊行動パターンの基礎分析 : 大阪キタ・ミナミ地区の比較. 第 34 回土木計画学研究発表会講演集 (CD-ROM), 2006.
- [72] T. Hirao, H. Isozaki, E. Maeda, and Y. Matsumoto. Extracting important sentences with support vector machines. In *Proc. 19th Int. Conf. on Computational Linguistics*, pp. 342–348, 2002.
- [73] 平尾努, 磯崎秀樹, 前田英作, 松本裕治. Support vector machine を用いた重要文抽出法. 情報処理学会論文誌, Vol. 44, No. 8, pp. 2230–2243, 2003.
- [74] 賀沢秀人, 平尾努, 前田英作. Order svm: 一般化順序統計量に基づく順位付け関数の推定. 電子情報通信学会論文誌 D-II, Vol. J86-D-II, No. 7, pp. 926–933, 2003.

- [75] 柳本豪一, 大松繁. カーネル法を用いた関連フィードバックによる興味抽出. 電気学会論文誌 C (電子・情報・システム部門誌), Vol. Vol.126, No. 3, pp. 395–400, 2006.
- [76] R. Herbrich, T. Graepel, P. Bollmann-Sdorra, and K. Obermayer. Learning a preference relation for information retrieval. In *Proceedings of the AAAI Workshop Text Categorization and Machine Learning*, 1998.
- [77] R. Herbrich, T. Graepel, and K. Obermayer. Large margin rank boundaries for ordinal regression. In *Advances in Large Margin Classifiers*, pp. 115–132. 2000.
- [78] T. Qin, T. Y. Liu, W. Lai, X. D. Zhang, D. S. Wang, and H. Li. Ranking with multiple hyperplanes. In *Proceedings of the 30th Annual International ACM SIGIR Conference*, 2007.
- [79] 大津展之. パターン認識における特徴抽出に関する数理的な研究. 電子技術総合研究所研究報告, Vol. 818, , 1981.
- [80] 『グルメ情報検索サイトぐるなび』. <http://www.gnavi.co.jp/> (2008-12-14 確認).
- [81] 『ぐるなび web サービス』. <http://api.gnavi.co.jp/api/service.htm> (2008-12-14 確認).
- [82] 小野田崇. サポートベクターマシンその仕組みと応用 - 分類手法の新展開 - . オペレーションズ・リサーチ学会誌 2001 年 5 月号, Vol. 46, No. 5, p. 224, 2001.
- [83] K. Järvelin and J. Kekäläinen. Ir evaluation methods for retrieving highly relevant documents. In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 41–48, 2000.

- [84] R. Agrawal and R. Srikant.: Mining sequential patterns. In *Proc. of ICDE1995*, pp. 3–14, 1995.
- [85] Jiawei Han and Micheline Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann, second edition edition, 2006.
- [86] 平手勇宇, 山名早人. 時間情報を含むシーケンシャルパターンマイニングの一般化. In *DEWS2006*, 2006.
- [87] J. Pei, J. Han, B. Mortazavi-Asl, H. Pinto, Q. Chen, U. Dayal, and C. M. Hsu. Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth. In *Proc. of ICDE2001*, pp. 215–224, 2001.
- [88] M. Zaki. An efficient algorithm for mining frequent sequences. In *Machine Learning*, Vol. 40, pp. 31–60, 2001.
- [89] Zhenglu Yang, Masaru Kitsuregawa, and Yitong Wang. Paid: Mining sequential patterns by passed item deduction in large databases. In *Database Engineering and Applications Symposium, International*, pp. 113–120, Los Alamitos, CA, USA, 2006. IEEE Computer Society.
- [90] Daniel S. Hirschberg. Algorithms for the longest common subsequence problem. *Journal of the ACM (JACM)*, Vol. 24, No. 4, pp. 664–675, oct 1977.
- [91] L. Bergroth, H. Hakonen, and T. Raita. A survey of longest common subsequence algorithms. In *Proc. 7th Internat. Symp. on String Processing Information Retrieval (SPIRE2000)*, pp. 39–48, Spain, 2000.
- [92] 宇根田純治, 横田治夫. Web ログの共通シーケンス解析. 電子情報通信学会技術研究報告. DE, データ工学, 102(64), pp. 7–12, 2002.
- [93] 戸田誠二. Lcs を用いたアクセスログ解析の並列処理による性能向上. 第 13 回データ工学ワークショップ論文集, 2004.

- [94] 石塚淳, 鈴木優, 川越恭二. 内容を考慮した移動軌跡データの類似検索手法.
第18回データ工学ワークショップ (DEWS2007) 論文集, E1-7, 2007.
- [95] 『ドコイク?』 . <http://www.doko.jp/> (2008-12-14 確認).

研究業績

A.査読付き論文誌

[A1] 奥 健太, 中島 伸介, 宮崎 純, 植村 俊亮, 加藤 博一: ``情報推薦におけるユーザの価値判断基準モデルに基づくコンテキスト依存型ランキング方式'', 情報処理学会論文誌: データベース, (TOD41), 2009年3月.

[A2] 奥 健太, 中島 伸介, 宮崎 純, 植村 俊亮: ``状況依存型ユーザ嗜好モデリングに基づく Context-Aware 情報推薦システム'', 情報処理学会論文誌: データベース, Vol.48, No.SIG 11 (TOD34), pp.162-176, 2007年6月.

B.Letters

[B1] 奥 健太, 中島 伸介, 宮崎 純, 植村 俊亮: ``Context-Aware SVM に基づく状況依存型情報推薦方式の提案'', 日本データベース学会論文誌 (DBSJ Letters), Vol.5, No.1, pp.5-8, 2006年6月.

C.著書 (分担執筆)

[C1] Kenta Oku, Shinsuke Nakajima, Jun Miyazaki, Shunsuke Uemura and Hirokazu Kato: ``Context-Aware Ranking Method for Information Recommendation" , Advances in Communication Systems and Electrical Engineering, Lecture Notes in Electrical Engineering, Vol.4, pp.319-337, Springer-Verlag, Mar. 2008.

D. 査読付き国際会議

[D1] Kenta OKU, Shinsuke NAKAJIMA, Jun MIYAZAKI, Shunsuke UEMURA, and Hirokazu KATO: "A Recommendation Method Considering User's Time Series Contexts", The Third International Conference on Ubiquitous Information Management and Communication (ICUIMC2009), pp.495-500, January 2009.

[D2] Kenta OKU, Shinsuke NAKAJIMA, Jun MIYAZAKI, Shunsuke UEMURA, and Hirokazu KATO: "A Ranking Method based on Users' Contexts for Information Recommendation", The Second International Conference on Ubiquitous Information Management and Communication (ICUIMC2008), pp.306-312, January 2008.

[D3] Kenta OKU, Shinsuke NAKAJIMA, Jun MIYAZAKI and Shunsuke UEMURA: "Investigation for Designing of Context-Aware Recommendation System Using SVM", The 2007 IAENG International Conference on Internet Computing and Web Services (IMECS2007), pp.970-975, March 2007.

[D4] Kenta Oku, Shinsuke Nakajima, Jun Miyazaki and Shunsuke Uemura: "Context-Aware SVM for Context-Dependent Information Recommendation", Proc. of International Workshop on Future Mobile and Ubiquitous Information Technologies (FMUIT2006), May 2006.

E. 査読なし国際会議

[E1] Kenta Oku, Shinsuke Nakajima, Jun Miyazaki and Shunsuke Uemura: "Context-Dependent Information Recommendation using SVM", Proc. of the 1st China-Kyoto Student Workshop on Digital Content and Web Computing (CKSW2006), pp.47-51, March 2006.

F.国内発表

[F1] 奥 健太, 中島 伸介, 宮崎 純, 植村 俊亮, 加藤 博一: ``ユーザコンテキストの時系列性を考慮した状況依存型情報推薦方式の有効性評価'', 第1回データ工学と情報マネジメントに関するフォーラム (DEIM2009), E5-4, 2009年3月.

[F2] 奥 健太, 中島 伸介, 宮崎 純, 植村 俊亮, 加藤 博一: ``ユーザの時系列コンテキストを考慮した情報推薦方式の提案'', 情報処理学会研究報告, Vol.2008, No.88, 2008-DBS-146, pp121-126, 2008年9月.

[F3] 奥 健太, 中島 伸介, 宮崎 純, 植村 俊亮, 加藤 博一: ``情報爆発時代に向けたコンテキストウェア情報推薦方式のための特徴パラメータ最適化手法'', 情報処理学会第70回全国大会論文集, pp.5-71~5-72, 2008年3月.

[F4] 奥 健太, 中島 伸介, 宮崎 純, 植村 俊亮, 加藤 博一: ``ユーザが重要視するパラメータの推定によるコンテキスト依存型ランキング方式'', 電子情報通信学会 第19回データ工学ワークショップ (DEWS2008) 論文集, B7-6, 2008年3月.

[F5] 奥 健太, 中島 伸介, 宮崎 純, 植村 俊亮, 加藤 博一: ``情報推薦におけるコンテキスト依存型ランキング方式の提案'', 電子情報通信学会技術研究報告, Vol.107, No.131, DE2007-22 ~ 114, pp.121-126 / 情報処理学会研究報告, Vol.2007, No.65, 2007-DBS-143, pp121-126, 2007年7月.

[F6] 奥 健太, 中島 伸介, 宮崎 純, 植村 俊亮: ``SVM を用いた状況依存型情報推薦方式のシステム構築に関する考察'', Proc. of DBWeb2006, 情報処理学会シンポジウムシリーズ Vol.2006, No.16, pp.51-58, 2006年11月.

[F7] 奥 健太, 中島 伸介, 宮崎 純, 植村 俊亮: ``Context-Aware SVM に基づく状況依存型情報推薦方式の提案'', 電子情報通信学会 第17回データ工学ワークショップ (DEWS2006) 論文集, 1A-oi2, 2006年3月.

G.招待講演

[G1] 奥 健太: ``ユーザコンテキストを考慮した情報推薦方式に関する研究", 第1回 jDB ワークショップ, 2008 年 6 月.

H.受賞

[H1] 2008 年 09 月 情報処理学会データベースシステム研究会 学生奨励賞 (iDB2008)

[H2] 2008 年 06 月 日本データベース学会 / 電子情報通信学会データ工学研究会 / 情報処理学会データベースシステム研究会 優秀若手研究者賞 (DBSJ / IEICE SIGDE / IPSJ SIGDBS Distinguished Young Researcher Award) (DEWS2008)

[H3] 2007 年 07 月 情報処理学会データベースシステム研究会 学生奨励賞 (DBWS2007)

[H4] 2007 年 03 月 Best Student Paper Award of The 2007 IAENG International Conference on Internet Computing and Web Services (IMECS2007)

[H5] 2007 年 03 月 Best Student Paper Award of The International MultiConference of Engineers and Computer Scientists 2007 (IMECS2007)