

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

奥 健太[†] 中島 伸介^{††} 宮崎 純[†] 植村 俊亮^{†††} 加藤 博一[†]

[†] 奈良先端科学技術大学院大学 情報科学研究科 〒630-0101 奈良県生駒市高山町 8916-5

^{††} 京都産業大学 コンピュータ理工学部 〒603-8555 京都市北区上賀茂本山

^{†††} 奈良産業大学 情報学部 情報学科 〒636-8503 奈良県生駒郡三郷町立野北 3-12-1

E-mail: [†]{kenta-o,miyazaki,kato}@is.naist.jp, ^{††}nakajima@cse.kyoto-su.ac.jp,

^{†††}uemurashunsuke@nara-su.ac.jp

あらまし 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、情報推薦においてユーザコンテキストを考慮することは重要な課題である。本研究では、ユーザコンテキストを「ユーザの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」と定義したうえで、コンテキスト依存型情報推薦方式の確立を目標としている。ユーザコンテキストには、ユーザが直面しているその時点での状況だけでなく、ユーザがその時点までに取ってきた行動やその時点以降に予定している行動といった、過去/未来への時系列的な要素も含まれる。我々の先行研究においては、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を提案しているが、本研究では、被験者実験を通じてこの提案方式の有効性の評価を行う。

キーワード 情報推薦, ユーザコンテキスト, 時系列コンテキスト, 行動パターン

Evaluation of effectiveness of a context-aware recommendation method considering users' action history

Kenta OKU[†], Shinsuke NAKAJIMA^{††}, Jun MIYAZAKI[†], Shunsuke UEMURA^{†††}, and Hirokazu KATO[†]

[†] Graduate School of Information Science, Nara Institute of Science and Technology,
8916-5 Takayama, Ikoma City, Nara, 630-0101, Japan

^{††} Kyoto Sangyo University, Faculty of Computer Science and Engineering
Motoyama, Kamigamo, Kita-Ku, Kyoto-City 603-8555, Japan

^{†††} Department of Informatics, Faculty of Informatics, Nara Sangyo University,
12-1, Tatsuno-kita 3-chome, Sango-cho, Ikoma-gun, Nara, 636-8503, Japan

E-mail: [†]{kenta-o,miyazaki,kato}@is.naist.jp, ^{††}nakajima@cse.kyoto-su.ac.jp,

^{†††}uemurashunsuke@nara-su.ac.jp

Abstract It is important to consider users' contexts in information recommendation, since users' preferences are changed according to their contexts. Our purpose is to develop a context-aware recommendation method based on users' contexts defined as the users' situations and conditions which influence the users' preferences in information choice. The users' contexts include not only users' current contexts but also contexts in the past and future such as users' action history. In our previous work, we proposed a context-aware recommendation method considering the users' action history. In this paper, we show effectiveness of the proposed method by experimental results.

Key words Recommendation Method, User's Contexts, Time Series Contexts, User's Behavior Pattern

1. はじめに

ユーザコンテキスト (ユーザの状況) に応じた情報をユーザに提供するコンテキスト依存型情報推薦技術が注目されている。我々は、これまでに、「情報選択過程におけるユーザの嗜好に影響を及ぼす、そのユーザを取り巻く状況」をユーザコンテキストと定義したうえで、このユーザコンテキストを考慮したコンテキスト依存型情報推薦方式 [1] を提案した。

我々の先行研究 [2] では、推薦時コンテキストおよび過去/未来コンテキストを総合的に考慮した情報推薦方式を提案した。ここで、推薦時コンテキストとは、ユーザの予算や同伴者など、ユーザが推薦システムから推薦情報を受け取る時点の状況のことをいう。また、過去コンテキストとはユーザがこれまでに取ってきた行動をいい、未来コンテキストとはユーザがこれから予定している行動をいう。

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、その時点までに取ってきた行動や、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。先行研究 [2] の推薦方式は、このことを考慮に入れた推薦を行うものである。

推薦処理は、推薦時コンテキストに依存する部分の処理と、過去/未来コンテキストに依存する部分の処理とを分けて、段階的に行う。

まず、ユーザの行動履歴データベースというものを考える。行動履歴データベースとは、これまでにユーザが取ってきた行動に関するデータを、そのときのコンテキスト情報と共に時系列に蓄積したものである。また、行動履歴データベースにおいて、頻出する一連の行動を行動パターンという。

先行研究の推薦方式では、ユーザの行動履歴データベースから、ユーザの推薦時コンテキストにおける行動パターンを抽出する。そして、その行動パターンとユーザの過去/未来コンテキストとを照合することで、ユーザが次に取りそうな行動を予測する。最終的に、予測された行動に基づいてユーザへの推薦情報を決定する。このような手順を踏むことにより、推薦時コンテキストおよび過去/未来コンテキストを総合的に考慮した情報推薦が可能となる。

先行研究においては、この推薦方式に関する有効性の検証を行っていなかったため、本稿では、被験者実験を通じて、我々の提案方式の有効性を検証する。

2. 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

2.1 概要

我々はこれまでに、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 [2] を提案した。ここでは、この推薦方式の概要について述べる。

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、その時点までに取ってきた行動や、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) 「一人でいるとき、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、“恋人といるとき”や“一人でいる”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。

我々の推薦方式 [2] では、現在の状況を推薦時コンテキストとして、その時点までに取ってきた行動を過去コンテキスト、その時点以降に予定している行動を未来コンテキストとして、それぞれ扱い、これらを総合的に考慮した情報推薦を行う。

推薦処理は、推薦時コンテキストに依存する部分の処理と、過去/未来コンテキストに依存する部分の処理とを分けて、段階的に行う。

まず、ユーザの行動履歴データベースというものを考える。行動履歴データベースは、これまでにユーザが取ってきた行動に関するデータを、そのときのコンテキスト情報と共に時系列に蓄積したものである。行動履歴データベースは、ユーザ一人一人に作られるものとする。ユーザ u の行動履歴データベースを Ω^u と表記する。

具体的には、以下の手順により行う。

- (1) 類似ユーザの選定による対象とする行動履歴データベースの特定
- (2) 推薦時コンテキストに関連する部分行動履歴データベースの特定
- (3) 推薦時コンテキストにおける行動パターンの抽出
- (4) 過去/未来コンテキスト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

最初の手順として、対象ユーザとの嗜好の類似度に基づいて M 人の類似ユーザ $v_1, v_2, \dots, v_N$ を選定し、その類似ユーザの行動履歴データベース $\Omega^{v_1}, \Omega^{v_2}, \dots, \Omega^{v_N}$ を参照する。つづいて、その行動履歴データベースから対象ユーザの推薦時コンテキストに関連する部分行動履歴データベース (Ω^* と表記する) を特定し、このデータベースから行動パターンを抽出する。ここで、行動パターンとは、行動履歴データベースにおいて頻出する一連の行動のことをいう。行動履歴データベースからの行動パターンの抽出には、シーケンシャルパターンマイニング手法 [3] や LCS (Longest Common Subsequence) 抽出手法 [4] などを用いる。

そして、抽出された行動パターンから、対象ユーザの過去/未来コンテキストに合った行動パターンを特定し、この行動パターンに沿って、ユーザが次に取りそうな行動を予測する。この行動の予測は、次節で説明する行動予測スコアに基づいて行われる。最終的に、この行動予測スコアに基づいて、推薦候補アイテムに対しランキングを行う。

表 1 推薦時コンテキストに関連した行動履歴データベース

Table 1 An action history database related to the current context.

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 2 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)

Table 2 A result of extracting frequent action patterns by a sequential pattern mining method (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

2.2 シーケンシャルパターンマイニング手法による行動予測スコアの算出

本節では、シーケンシャルパターンマイニング手法 [3] (以降、SPM と表記する) による行動予測スコアの算出方法について述べる。

ここでは、既に推薦時コンテキストに関連する部分行動履歴データベース Ω^* として、表 1 に示す行動履歴データベースが特定されているとする。表 1 の部分行動履歴データベースに対し、SPM を適用すると、表 2 のような結果が得られる。ここで、表 2 中の count は対応する行動シーケンスの出現回数、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、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。

$$\text{sup}(\langle a, b \rangle) = \frac{n(\langle a, b \rangle)}{N} \quad (1)$$

ここで、 $n(\langle a, b \rangle)$ は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。

また、シーケンス $\langle a, b \rangle$ のコンフィデンス値 $\text{conf}(\langle a, b \rangle)$

表 3 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

Table 3 A result of choosing frequent action patterns matching for the user's past context by a sequential pattern mining method.

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

は、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。

$$\text{conf}(\langle a, b \rangle) = \frac{\text{sup}(\langle a, b \rangle)}{\text{sup}(\langle a \rangle)} \quad (2)$$

$\text{conf}(\langle a, b \rangle)$ 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 2 の例では、minsup = 0.2 としたときの抽出結果を示している。

いま、対象ユーザの過去コンテキストが、

$$T = \begin{cases} \text{"買物する"} \text{ at "17:00"} \\ \text{"食事する"} \text{ at "18:00"} \end{cases} \quad (3)$$

であったとする。このとき、過去コンテキストに合った行動パターンとして、表 3 に示す行動パターンが特定される。

さらに、表 3 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 4 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。

なお、この行動予測スコアは、対象ユーザの推薦時コンテキストに関連するとして特定された部分行動履歴データベースに基づいて算出されたものである。したがって、この行動予測スコアは、推薦時コンテキストおよび過去 (未来) コンテキストの両方を考慮したスコアを意味している。

また、我々の先行研究 [1] 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。本論文においても、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、我々の先行研究の推薦方式 [1] により行う。本稿での提案方式は、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。

表 4 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

Table 4 Score for each action by a sequential pattern mining method.

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

3. 評価実験

2. 章で述べた推薦方式 (以降, 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.

3.1 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?』[5] 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 5 に示す施設ジャンルに属するデータを選んだ. なお, 表 5 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 5 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 5 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 5 に示すとおりである (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.

3.2 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.

3.2.1 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) を想定してもらう. 提示するコンテキストは, 表 6 に示す推薦時コン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.

3.2.2 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) $\text{sim}_a^{\text{user}}$: コンテキストを考慮しないユーザ類似度

(b) $\text{sim}_b^{\text{user}}$: 推薦時コンテキストのみを考慮したユーザ類似度

(c) $\text{sim}_c^{\text{user}}$: 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 $\text{sim}^{\text{user}}(A, B)$ は, 次式により表される.

$$\text{sim}^{\text{user}}(A, B) = \frac{\sum(e_{Ai} - \bar{e}_A)(e_{Bi} - \bar{e}_B)}{\sqrt{\sum(e_{Ai} - \bar{e}_A)^2} \sqrt{\sum(e_{Bi} - \bar{e}_B)^2}} \quad (4)$$

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, $\text{sim}_a^{\text{user}}(A, B)$ 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. $\text{sim}_b^{\text{user}}(A, B)$ 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, $\text{sim}_c^{\text{user}}(A, B)$ 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.

3.2.3 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(3.2.1 項) 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

表 5 実験で用いた施設データ

Table 5 Facility data used for experiments.

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

表 6 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

Table 6 Users' current contexts and past contexts used for experiments.

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”)」

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 A を除く被験者の全データ

(IV) Ω_{IV} : 被験者 A を除く被験者のデータのうち、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。

3.2.4 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(3.2.3 項) 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(3.2.1 項) 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 (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、2. 章で説明した、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) 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) による行動予測スコアを $Score_i^{LCS}$ と表記する。

3.2.5 フェーズ 5: スコア算出結果に対する評価

フェーズ 4(3.2.4 項) 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。

本実験では、評価尺度として、 R 適合率 [6] 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。

$$\text{適合率} = \frac{\text{ランキング結果に含まれる適合データの数}}{\text{ランキング結果に含まれるデータの数}} \quad (5)$$

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。

3.3 実験結果および考察

3.3.1 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 1 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 1 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。

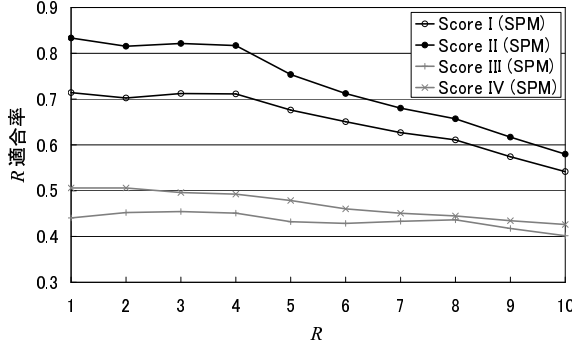


図 1 推薦時コンテキスト考慮時/非考慮時における SPM によるランキング結果に対する R 適合率

Fig. 1 R precision of ranking by an SPM when considering / not considering users' current context.

同様に、LCS による結果を図 2 に示す。なお、LCS におけるパラメタは、 $\text{min.Count} = 1$, $\text{min.Length} = 2$ とした。図 2 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。

また、図 1 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。

なお、図 1 および図 2 において、 $R = 4$ までの R 適合率がほぼ水平になった。この理由について説明する。まず、評価用のデータはコンテキストごとに 20 件であり、3.2.1 項で説明したように、各施設カテゴリにつき 4 件ずつの施設データにより構成される。行動予測スコアも施設カテゴリに対応して算出されるため、この行動予測スコアに基づいたランキング結果も、同一カテゴリのデータが 4 件ずつ連続したものとなる。被験者も、同一のカテゴリに対しては、ほぼ同じ評価を与えている。特に上位 4 件のデータに対しては、適合データとなる場合が多かったため、結果として、上位 4 件に対する R 適合率がほぼ一定となった。

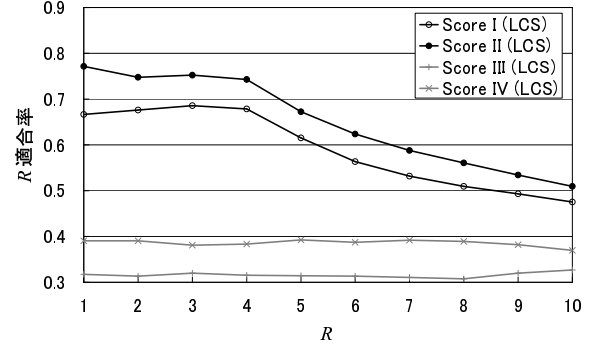


図 2 推薦時コンテキスト考慮時/非考慮時における LCS によるランキング結果に対する R 適合率

Fig. 2 R precision of ranking by an LCS when considering / not considering users' current context.

3.3.2 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 3 は、SPM による行動予測スコア $\text{Score}_{IV}^{\text{SPM}}$, $\text{Score}_V^{\text{SPM}}$, $\text{Score}_{VI}^{\text{SPM}}$, $\text{Score}_{VII}^{\text{SPM}}$ 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。

$\text{Score}_{IV}^{\text{SPM}}$ 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、 $\text{Score}_V^{\text{SPM}}$, $\text{Score}_{VI}^{\text{SPM}}$, $\text{Score}_{VII}^{\text{SPM}}$ は、それぞれ、3.2.2 項で述べた、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 4 に示す。

図 3 および図 4 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証実験により検証していく必要がある。

4. ま と め

本稿では、我々の提案方式である、推薦時コンテキストおよび過去/未来コンテキストを総合的に考慮した情報推薦方式の有効性の検証を行った。

被験者実験を通じて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。

今後は、行動パターン抽出手法として適用した、シーケンシャルパターンマイニング手法や LCS 抽出手法などにおける、パラメタ調整についても検討する。

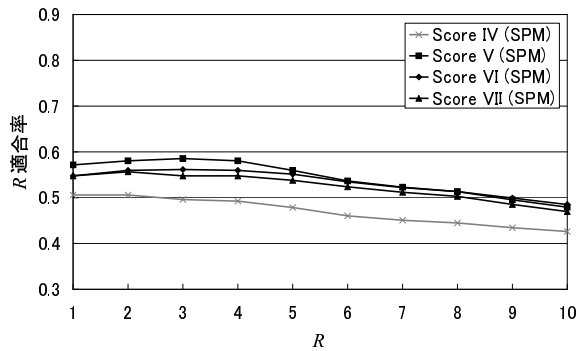


図 3 類似ユーザを考慮したときの SPM によるランキング結果に対する R 適合率

Fig. 3 R precision of ranking by an SPM when considering similar users.

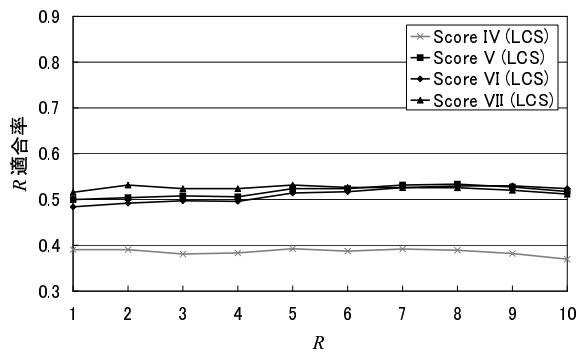


図 4 類似ユーザを考慮したときの LCS によるランキング結果に対する R 適合率

Fig. 4 R precision of ranking by an LCS when considering similar users.

謝 辞

本研究の一部は、文部科学省科学研究費補助金特定領域研究「情報爆発時代に向けた新しい IT 基盤技術の研究」(A01-34, 課題番号 19024058), 文部科学省科学研究費補助金若手研究(B) (課題番号: 20700089) による。ここに記して謝意を表します。

文 献

- [1] 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.
- [2] 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.
- [3] R. Agrawal and R. Srikant: “Mining Sequential Patterns,” In Proc. of ICDE1995, pp.3-14, 1995.
- [4] 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, Vol.48, No.SIG 11 (TOD34), pp.38-48, 2007 年 6 月.

- [5] 『ドコイク?』, <http://www.doko.jp/> (2008-12-14 確認).
- [6] 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 (2002).