

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

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 過去/未来に起こった/起こりうる状況、すなわち時系列コンテキストを考慮した情報推薦方式を提案する。時系列コンテキストを考慮した情報推薦として、ユーザの時系列的な行動パターンに沿った情報を提供する情報推薦技術がある。これは対象とするユーザの行動パターンと類似する行動パターンを持つユーザの情報を参照し、推薦情報を判定しようというものである。しかし、ユーザの行動パターンは、そのときのユーザコンテキストにより変化するため、推薦時のユーザコンテキストに限定した行動パターンを参照する必要がある。本稿では、ユーザの時系列的な行動パターンがその時々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、その上で、過去/未来にユーザが取った/取りうる行動がユーザの現在の価値判断に及ぼす影響を考慮した情報推薦方式を提案する。

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

A Recommendation Method Considering User's Time Series Contexts

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 This paper proposes a recommendation method considering user's time series contexts which are situation occurred/occurring in the past/future. There are some recommendation methods which provide information suitable for users' time series behavior patterns as the recommendation methods considering them. These methods provide information referring to the other users which have a similar behavior pattern to an active user's pattern. However, since a user's behavior pattern changes depending on the user's contexts, the methods need to refer to the other users' behavior patterns related to the current user's contexts. In this paper, we propose a recommendation method considering the user's time series contexts considering that user's behavior pattern changes depending on the user's contexts.

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

1. はじめに

インターネット技術等の発展に伴い、我々がアクセスできる情報が膨大なものになってきた [1]。この膨大な情報の中から、

ユーザの嗜好に合った情報を提供する情報推薦技術が注目されている。また、近年のモバイル技術の発展により、我々は、「いつでも」「どこでも」、日常的に情報システムを利用することができるようになった。こうした中、ユーザの状況 (コンテキス

トと呼ぶ)を考慮したコンテキストウェア (Context-Aware) システムに関する研究が盛んに行われている [2]. 中でも、近年、ユーザのコンテキストを考慮した情報推薦技術、すなわちコンテキストウェア情報推薦技術に関する研究も注目されるようになってきた [3].

ユーザのコンテキストのうち、位置情報に基づいた旅行情報を提供するシステムとして、Cheverst らが提案している GUIDE [4], Setten らが提案している COMPASS [5] 等がある. ユーザの位置情報は GPS 等から容易に取得できるため、これまでもユーザの位置情報に基づいた情報推薦技術に関する研究は数多く行われている. しかし、ユーザの価値判断は単純にユーザの位置情報だけに影響を受けるものではなく、他にもさまざまな状況に影響を受けるものであると考えられる. 例えば、飲食店を推薦対象とすれば、ユーザが要求する飲食店は、ユーザの予算や同伴者、その日が休日であるか否か、天気や気温等、状況によって大きく変化するものである. したがって、これら種々の状況を総合的に考慮した上で、推薦情報を判定すべきである. コンテキストウェア情報推薦技術を確立するためには、ユーザの価値判断に影響を及ぼす状況というものを考慮に入れなければならない. そこで、我々はこれまでに、“情報選択に対するユーザの価値判断に影響を及ぼすユーザを取り巻く状況”を「ユーザコンテキスト」と定義し、このユーザコンテキストを考慮した情報推薦方式を提案した [3].

ただし、これまでの研究で扱ってきたユーザコンテキストは、先に挙げたような、天気や気温、ユーザの予算、同伴者等、推薦時のユーザの状況に限定したものであった. “ユーザの価値判断に影響を及ぼす状況”というものを広い意味で捉えると、推薦時の状況だけでなく、過去に起こった状況、さらには未来に起こりうる状況も、ユーザコンテキストとして含めるべきである. 例えば、「居酒屋で飲んだ後はカラオケで歌いたい」、「昼間に野球の試合をした後は焼肉屋で皆で打上げをしたい」、「1 時間後に友達と駅で待合せをしているので、近くのカフェで時間をつぶしたい。」等といった場面が考えられる. この例のように、「居酒屋で飲んだ」、「昼間に野球の試合をした」、「1 時間後に友達と駅で待合せをする」といった過去もしくは未来の状況は、ユーザの現在の価値判断に影響を及ぼすものであるといえる. そこで、本稿では、“ユーザの現在の価値判断に影響を及ぼす、過去/未来に起こった/起こりうる状況”を「時系列コンテキスト」と定義し、この時系列コンテキストを考慮した情報推薦方式を提案する. 特に、今回は問題を簡略化するため、過去/未来に起こった/起こりうる状況を過去/未来にユーザが取った/取りうる行動に限定して扱うことにする.

時系列コンテキストをこのようにユーザの行動ということに限定すると、時系列コンテキストを考慮した情報推薦の一つとして、ユーザの時系列的な行動パターンに沿った情報を推薦することが挙げられる. 例えば、あるユーザが「居酒屋で飲んだ後にカラオケで歌う。」といった行動パターンを持っていたとすると、このユーザが居酒屋に行った後には、カラオケに行く可能性が高いと推測されるため、近くのカラオケ店を推薦する等である. 篠田らは、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 [6]. 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.

しかしながら、ユーザの行動パターンもやはりその時のユーザコンテキストによって変化するものと考えられる. 例えば、同じく居酒屋で飲んだ後であったとしても、「友達いる場合には、次はカラオケで歌いたい。」が、「恋人がいる場合には、次はバーで飲みたい。」といった場面が考えられる. したがって、単にユーザの行動履歴の類似度だけでなく、ユーザコンテキストごとの行動パターンの類似度を考慮に入れる必要がある. そこで本稿で提案する時系列コンテキストを考慮した情報推薦方式では、ユーザの行動パターンがその時々の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、その上で、過去/未来にユーザが取った/取りうる行動がユーザの現在の価値判断に及ぼす影響を考慮した情報推薦方式を提案する.

2. 関連研究

2.1 ユーザの時系列行動パターンに沿った情報推薦技術

ユーザの時系列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.

山元らは、Web アクセスログを解析することで頻出アクセスパターンを発見し、そのパターンに沿った Web ページ推薦手法を提案している [7]. 例えば、Web ページ A, B, C, D, E に対し、 $\langle A, B, C, D, E \rangle$ という頻出アクセスパターンが発見されたとする. このとき、対象ユーザが $\langle A, B, C \rangle$ と順に Web ページを閲覧すると、つづく、Web ページ D, E が推薦候補ページとして加えられる.

石塚らは、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 [8]. 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.

篠田らは、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 [6]. 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.

以上のように、ユーザの時系列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、ユーザのアクセスログや行動履歴等からユーザが取りうる行動パターンを抽出し、その行動パターンに基づいて何らかの情報を提供しようというものである. しかしながら、ユーザの行動パターンもやはりその時のユーザコンテキストによって変化するものと考えられる. したがって、単にユーザの行動履歴の類似度だけでなく、ユーザコンテキストごとの行動パターンの類似度を考慮に入れる必要がある.

本稿で提案する時系列コンテキストを考慮した情報推薦方式では、ユーザの行動パターンがその時々の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、その上で、過去/未来に

ユーザが取った/取りうる行動がユーザの現在の価値判断に及ぼす影響を考慮した情報推薦方式を提案する。

2.2 時系列行動パターン抽出手法

時系列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 [9] や隠れマルコフモデル (HMM) [10], LCS (Longest Common Subsequences) 抽出手法 [7] 等が用いられる。

ここでは、シーケンシャルパターンマイニングについて簡単に説明する。シーケンシャルパターンマイニングは、シーケンスデータベースから頻出シーケンスを抽出する手法である [9]。シーケンスとは、順序を持つアイテムの列であり、 $\langle a, b \rangle$ 等のように記述する。 $\langle a, b \rangle$ は、アイテム a が発生した後に、アイテム b が発生したことを表す。シーケンス s がシーケンスデータベース内において出現する頻度のことをシーケンス s のサポート値 (support) という。このサポート値が、定義された最小サポート値 (minimum support) を超えるようなシーケンスが頻出シーケンシャルパターンとして抽出される。シーケンシャルパターンマイニングをユーザの行動履歴データベースに対して適用することにより、ユーザの頻出行動パターンを抽出することができる。

3. ユーザコンテキストの定義

ユーザコンテキストとは、一般にユーザを取り巻くさまざまな状況を意味するが、我々は独自に、ユーザコンテキストを“情報選択に対するユーザの価値判断に影響を及ぼすユーザを取り巻く状況”と定義している [3]。

この定義に則ったユーザコンテキストとして、例えば以下に示すものが挙げられる。

- 時間情報：季節，曜日，時刻，...
- 気象情報：天気，気温，湿度，...
- ユーザ情報：休日/非休日，予算，余裕時間，目的，気分，体調，...
- 同伴者情報：関係（家族や恋人，友人，上司・部下等），人数，...
- 周辺環境：現在地のエリアタイプ（繁華街や郊外，観光地等），周辺施設の分布状況，混雑状況，...

どのコンテキストがユーザの価値判断に影響を及ぼすのかといったことは、推薦対象とするコンテンツに依存するため、推薦対象コンテンツに基づいたユーザコンテキストの設計が必要である。我々は、これまでの研究において、上記のうち時間情報および気象情報，ユーザ情報，同伴者情報をユーザコンテキストとして扱った情報推薦方式を提案し、その有効性を示した [3]。

ただし、これまでの研究は推薦時のユーザコンテキストを扱ったものであった。上記に挙げたユーザコンテキストは、推薦時だけでなく、そのときを基準に過去さらには未来へと時間軸方向に拡張される。ユーザの価値判断は、推薦時の状況だけでなく、過去や未来の状況にも影響を受けるものと考えられるからである。そこで我々は、“現在のユーザの価値判断に影響を及ぼす、過去/未来に起こった/起こりうる状況”を「時系列コンテキスト」と定義し、推薦時のコンテキストに加え、時

表 1 ユーザの行動履歴データベース (例)

SID	データID	アイテム利用日時	利用アイテム	利用時のユーザコンテキスト		
				時刻	休日/非休日	予算(円) 同伴者
1	1	2008/8/1 12:00	a	1200	非休日	3,000 なし
	2	2008/8/1 15:00	b	1500	非休日	3,000 なし
2	3	2008/8/2 13:00	c	1300	休日	5,000 友人
	4	2008/8/2 16:00	d	1600	休日	5,000 友人
	5	2008/8/2 18:00	e	1800	休日	5,000 友人

系列コンテキストを考慮した情報推薦方式の確立を目指す。特に、本稿では、過去/未来に起こった/起こりうる状況を過去/未来にユーザが取った/取りうる行動に限定する。つまり、時系列コンテキストを“現在のユーザの価値判断に影響を及ぼす、過去/未来にユーザが取った/取りうる行動”と置き換え、問題を簡略化する。これは、あらゆる状況が考えられる時系列コンテキストのうち、ユーザの行動が最も直接的にユーザの価値判断に影響を及ぼすという仮定に基づくものである。

次章以降では、過去/未来のユーザの行動を時系列コンテキストとして扱ったときの、情報推薦方式を提案する。

4. ユーザの行動履歴データベース

本稿では、ユーザの行動をユーザが利用した/するアイテムによって表現する。したがって、ユーザの行動履歴はユーザの利用アイテムを時系列に記録した情報として表される。表 1 に、ユーザの行動履歴データベースの例を示す。ユーザの行動履歴データベース内の 1 データには以下の情報が記録されている。

- セッション ID
- データ ID
- ユーザがアイテムを利用した日時
- ユーザが利用したアイテムの情報
- アイテム利用時のユーザコンテキストの情報

ここで、ユーザコンテキストは、「時刻」や「休日/非休日」、「予算」、「同伴者」等の状況を表す各要素を一つ一つの次元としたときの多次元の特徴ベクトルで表現される。例えば、ユーザコンテキストの特徴ベクトル C の各次元 c_i を次のように定義する。

$$\begin{aligned}
 c_1 &: \text{時刻} \quad (0 \leq c_1 \leq 1) \\
 c_2 &: \text{休日/非休日} \quad (c_2 \in \{0, 1\}) \\
 c_3 &: \text{予算} \quad (0 \leq c_3 \leq 1) \\
 c_4 &: \text{同伴者} = \text{なし} \quad (c_4 \in \{0, 1\}) \\
 c_5 &: \text{同伴者} = \text{家族} \quad (c_5 \in \{0, 1\}) \\
 c_6 &: \text{同伴者} = \text{友人} \quad (c_6 \in \{0, 1\})
 \end{aligned} \tag{1}$$

すると、表 1 の行動履歴データベース中の第 1 番目のデータにおけるユーザコンテキスト C_1 は次式のように表される。

$$C_1 = \{c_1, c_2, c_3, c_4, c_5, c_6\} = \{0.5, 0, 0.3, 1, 0, 0\} \tag{2}$$

また、ユーザの利用アイテムは次のいずれかで表現される。

- アイテムの固有 ID
- アイテムのジャンル
- アイテムの特徴を表現する特徴ベクトル

いずれの方法によりアイテムを表現するのが適切であるかは、推薦対象とするコンテンツに依存する。例えば、奈良の観光施設を対象コンテンツとした場合、「東大寺」や「興福寺」、「春日大社」、「円興寺」等は、固有の施設として意味があるため、特徴化せずにそれぞれ固有のアイテムとして表現する方が都合が良い。一方、日常的に利用するような飲食店等を対象コンテンツとした場合は、その店がどのような雰囲気なのか、利用料金はいくらくらいなのか等というように、アイテムを特徴化して表現した方が都合が良い場合がある。特に、「和食」や「カフェ」、「居酒屋」等といった、そのアイテムのジャンルに強い意味があるようなコンテンツの場合には、ジャンルの情報だけで表現する方法も考えられる。

5. 提案方式

ユーザの時系列コンテキストを考慮した情報推薦方式を提案する。本稿では、先述したとおり、時系列コンテキストをユーザが過去/未来に取った/取りうる行動とし、その行動をユーザの利用アイテムにより表現する。今回は説明のため、アイテムは $\{a, b, c, d, e\}$ の固有 ID により表現されるものとする。

いま、対象ユーザ A が表 2 に示すように、[時刻 = 12:00], [休日/非休日 = 休日], [予算 = 3,000], [同伴者 = 友人] というコンテキストに現在置かれているとする。また、ユーザ A の直近の行動履歴として、[時刻 = 9:00] においてアイテム a を、[時刻 = 11:00] においてアイテム b を連続的に利用していたとする。このとき、ユーザ A がアイテム a, b を利用したという行動を取ったことが、ユーザ A の現在の価値判断に影響を及ぼすと考え、アイテム a, b を利用したという行動をユーザ A にとっての時系列コンテキストとして扱う。提案の推薦方式は、この時系列コンテキストによるユーザ A の現在の価値判断への影響を考慮しながら、ユーザ A への推薦アイテムを判定するものである。

提案方式では、対象ユーザと嗜好が類似した他のユーザの情報を用いて推薦を行う、すなわち協調フィルタリング[1]の概念を用いて対象ユーザへの推薦アイテムを判定する。単純には、対象ユーザ A の直近の行動履歴 ($a \rightarrow b \rightarrow ?$) と照らし合わせながら、これと類似の行動パターンを取っている他のユーザを探し出す。そして、そのユーザの行動パターンを参照することで、対象ユーザ A への推薦アイテムを判定するものである。ここで、ユーザの行動パターンを $\langle a, b \rangle$ のように表す。

表 3 は、例としてユーザ A とは別のユーザであるユーザ B およびユーザ C、ユーザ D、それぞれの行動履歴データベースを示している。それぞれ、セッション ID、データ ID、アイテム利用日時、利用アイテム ID、利用時のユーザコンテキストの情報を持つ。この行動履歴データベースのうち、各ユーザの 1 日単位の行動を 1 セッションと定義する。ここで、各行動履歴データベースに対し、2.2 節で述べたシーケンシャルパターンマイニングを適用すると、行動パターン $\langle a, b, c \rangle$ および $\langle a, b, d \rangle$ が頻出パターンとして抽出される。したがって、この場合、アイテム c, d をユーザ A への推薦アイテムとして判定できそうである。

表 2 ユーザ A の直近の行動履歴

SID	データID	アイテム利用日時	利用アイテムID	利用時のユーザコンテキスト			
				時刻	休日/非休日	予算(円)	同伴者
1	1	2008/8/22 9:00	a	9:00	休日	3,000	友人
	2	2008/8/22 11:00	b	11:00	休日	3,000	友人
	現在	2008/8/22 12:00	?	12:00	休日	3,000	友人

表 3 他ユーザの行動履歴データベース

(a) ユーザBの行動履歴データベース

SID	データID	アイテム利用日時	利用アイテムID	利用時のユーザコンテキスト				コンテキストの類似度
				時刻	休日/非休日	予算(円)	同伴者	
1	1	2008/8/2 11:00	a	11:00	休日	2,000	友人	0.998
	2	2008/8/2 12:00	b	12:00	休日	2,000	友人	0.998
	3	2008/8/2 14:00	c	14:00	休日	2,000	友人	0.998
2	4	2008/8/5 18:00	d	18:00	非休日	1,000	なし	0.211
	5	2008/8/5 19:00	e	19:00	非休日	1,000	なし	0.218
	6	2008/8/16 10:00	a	10:00	休日	3,000	友人	0.999
3	7	2008/8/16 12:00	b	12:00	休日	3,000	友人	1.000
	8	2008/8/16 13:00	c	13:00	休日	3,000	友人	1.000
	9	2008/8/20 17:00	a	17:00	非休日	2,000	なし	0.218
4	10	2008/8/20 18:00	b	18:00	非休日	2,000	なし	0.225
	11	2008/8/20 20:00	d	20:00	非休日	2,000	なし	0.237
	12	2008/8/21 17:00	a	17:00	非休日	1,000	なし	0.204
5	13	2008/8/21 19:00	b	19:00	非休日	1,000	なし	0.218
	14	2008/8/21 20:00	d	20:00	非休日	1,000	なし	0.224
	15	2008/8/21 21:00	e	21:00	非休日	1,000	なし	0.229

(b) ユーザCの行動履歴データベース

SID	データID	アイテム利用日時	利用アイテムID	利用時のユーザコンテキスト				コンテキストの類似度
				時刻	休日/非休日	予算(円)	同伴者	
1	1	2008/8/5 11:00	c	11:00	休日	3,000	友人	1.000
	2	2008/8/3 12:00	c	12:00	休日	2,000	友人	1.000
2	3	2008/8/5 19:00	b	19:00	非休日	2,000	なし	0.231
	4	2008/8/5 20:00	d	20:00	非休日	2,000	なし	0.237
3	5	2008/8/10 12:00	a	12:00	休日	2,000	友人	0.998
	6	2008/8/10 14:00	e	14:00	休日	2,000	友人	0.996
4	7	2008/8/15 19:00	b	19:00	非休日	2,000	なし	0.231
	8	2008/8/15 20:00	c	20:00	非休日	2,000	なし	0.237
5	9	2008/8/17 11:00	b	11:00	休日	2,000	友人	0.998
	10	2008/8/17 12:00	a	12:00	休日	2,000	友人	0.998
	11	2008/8/17 14:00	c	14:00	休日	2,000	友人	0.998

(c) ユーザDの行動履歴データベース

SID	データID	アイテム利用日時	利用アイテムID	利用時のユーザコンテキスト				コンテキストの類似度
				時刻	休日/非休日	予算(円)	同伴者	
1	1	2008/8/2 12:00	a	12:00	休日	3,000	友人	1.000
	2	2008/8/2 13:00	a	13:00	休日	3,000	友人	1.000
	3	2008/8/5 10:00	a	10:00	休日	2,000	友人	0.997
	4	2008/8/9 11:00	e	11:00	休日	2,000	友人	0.998
	5	2008/8/9 12:00	a	12:00	休日	2,000	友人	0.998
3	6	2008/8/16 11:00	a	11:00	休日	3,000	友人	1.000
	7	2008/8/16 13:00	a	13:00	休日	3,000	友人	1.000
	8	2008/8/16 17:00	b	17:00	非休日	1,000	なし	0.204
4	9	2008/8/18 18:00	b	18:00	非休日	1,000	なし	0.211
	10	2008/8/18 19:00	e	19:00	非休日	1,000	なし	0.218
	11	2008/8/18 20:00	d	20:00	非休日	1,000	なし	0.224
5	12	2008/8/20 18:00	d	18:00	非休日	2,000	なし	0.225
	13	2008/8/20 19:00	e	19:00	非休日	2,000	なし	0.231

ただし、この方法は、「同じ行動パターンを持つユーザが常に同じ行動を取る」という仮定に基づくものであり、必ずしもこの仮定に当てはまるとは限らない。例えば、居酒屋で飲んだ後であったとしても、「友達という場合には、次はカラオケで歌いたい。」が、「恋人という場合には、次はバーで飲みたい。」といったように、ユーザの行動パターンもそのときのユーザコンテキストに依存するものであると考えられる。したがって、単にユーザの全体的な行動パターンを照らし合わせるのではなく、推薦時のユーザコンテキストに限定した行動パターンを参照する必要がある。

そこで、提案方式では、以下の手順により対象ユーザへの推薦アイテムを判定する。

- (1) 他ユーザの行動履歴データベースから、推薦時のユーザコンテキストに限定した部分データベースを抽出する。
 - (2) 対象ユーザの直近の行動履歴に照らし合わせたとき、他ユーザの行動履歴データベースから想定される行動パターンを抽出する。
 - (3) 抽出された行動パターンの頻度情報に基づき、アイテムに対するスコアを算出する。
- 以降、各手順について説明する。

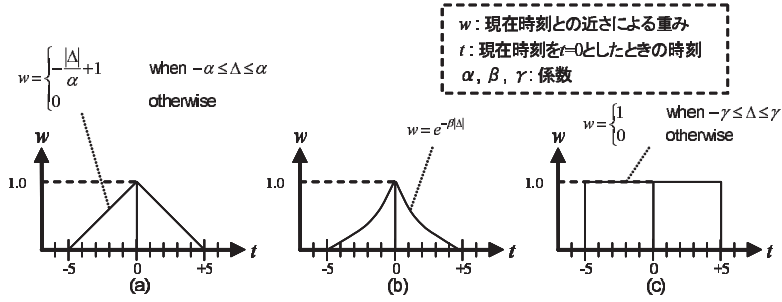


図1 現在時刻との近さによる重みの定義

表4 各行動パターンをサポート値

行動パターン	サポート値	行動パターン	サポート値	行動パターン	サポート値
$\langle a, a \rangle$	1	$\langle b, a \rangle$	1	$\langle (a, b), a \rangle$	0
$\langle a, b \rangle$	2	$\langle b, b \rangle$	0	$\langle (a, b), b \rangle$	0
$\langle a, c \rangle$	0	$\langle b, c \rangle$	0	$\langle (a, b), c \rangle$	3
$\langle a, d \rangle$	0	$\langle b, d \rangle$	0	$\langle (a, b), d \rangle$	0
$\langle a, e \rangle$	4	$\langle b, e \rangle$	0	$\langle (a, b), e \rangle$	0

5.1 推薦時のユーザコンテキストに限定した部分データベースの抽出

ユーザコンテキストに依存した行動パターンを参照するため、他ユーザの行動履歴データベースから、対象ユーザ A が現在置かれているコンテキストに限定した部分データベースを抽出する。

対象ユーザ A が置かれているコンテキスト C_A を、次式のように特徴ベクトルで表現する。

$$C_A = (0.5, 1, 0.3, 0, 0, 1) \quad (3)$$

このとき、他のユーザ u の行動履歴データベース中の第 i 番目のデータにおけるコンテキスト C_{ui} とのベクトルの類似度を求める。ベクトル間の類似度の定義は、コサイン尺度や内積等、さまざまなものが挙げられる [11] が、ここでは一般によく用いられているコサイン尺度の式を示す。

$$\cos(C_A, C_{ui}) = \frac{C_A \cdot C_{ui}}{\|C_A\| \|C_{ui}\|} \quad (4)$$

式 (4) より求められた類似度を、表 3 中の「コンテキストの類似度」列に示す。この類似度がある閾値以上のデータ集合を部分データベースとして抽出する。ここでは、閾値を 0.5 とした。これより、ユーザ B の行動履歴データベースからデータ ID={1, 2, 3, 6, 7, 8}、ユーザ C の行動履歴データベースからデータ ID={1, 2, 5, 6, 9, 10, 11}、ユーザ D の行動履歴データベースからデータ ID={1, 2, 3, 4, 5, 6, 7}、をそれぞれ持つデータ集合が部分データベースとして扱われる。以降、抽出された部分データベースを P と表記する。

5.2 想定される行動パターンの抽出

対象ユーザ A の直近の行動履歴が $\langle a, b \rangle$ であるので、前節にて抽出した部分データベース P より、 $\langle a, b, x \rangle$ ($x \in \{a, b, c, d, e\}$) が頻出行動パターンとなる場合、アイテム x を推薦アイテムとして判定することができる。

しかし、ユーザが $\langle a, b \rangle$ という行動パターンを取っていたとしても、実際には、ユーザの価値判断は必ずしもアイテム a, b の両方を利用したことに影響を受けているとは限らない。アイテム a だけを利用したことにより影響を受ける場合もあれば、アイテム b だけを利用したことにより影響を受ける場合もあると考えられる。したがって、ユーザ A への推薦アイテムを判定する際には、ユーザ A への価値判断への、アイテム a だけを利用したことによる影響、アイテム b だけを利用したことによる影響、アイテム a, b の両方を利用したことによる影響、それぞれを考慮に入れる必要がある。

行動履歴データベースからの頻出行動パターンの抽出は、2.2 節で取り上げた手法により行うことができる。いずれの手法が適切であるかは実験により検証する必要があるが、今回は、シーケンシャルパターンマイニングを用いることにする。対象ユーザ A の直近の行動履歴と照らし合わせたとき、想定されるすべての行動パターンは、 $\langle a, x \rangle, \langle b, x \rangle, \langle (a, b), x \rangle$ となる。ここで、 $\langle (a, b), x \rangle$ 中の (a, b) は、 a と b が共に発生することを意味し、その発生順序は問わない。また、 $\langle c, x \rangle, \langle d, x \rangle, \langle e, x \rangle$ 等の行動パターンについては、ユーザが直近に取っている行動履歴にアイテム a, b を利用したという情報しか含まれていないため、考慮する必要はない。

ここで、シーケンシャルパターンマイニングを用い、想定される各行動パターンをサポート値を部分データベース P から求める。部分データベース P にシーケンシャルパターンマイニングを適用し、各行動パターンをサポート値を求めた結果を表 4 に示す。この結果、行動パターン $\langle (a, b), c \rangle$ のサポート値が 3 と高かったため、アイテム c が推薦アイテムとして有力であることが分かる。また、行動パターン $\langle a, e \rangle$ のサポート値も 4 と高かった。これは、アイテム b を利用したか否かにかかわらず、アイテム a を利用したという行動に影響を受けて、アイテム e を利用していることを示している。

5.3 アイテムに対するスコアの算出

各行動パターンをサポート値に基づいて、アイテムに対するスコアを算出する。

今回の例の場合、アイテム x に対するスコア S_x は次式により求められる。

表5 別のコンテキストにおける各行動パターンのサポート値

行動パターン	サポート値	行動パターン	サポート値	行動パターン	サポート値
$\langle a, a \rangle$	0	$\langle b, a \rangle$	0	$\langle a, b, a \rangle$	0
$\langle a, b \rangle$	3	$\langle b, b \rangle$	0	$\langle a, b, b \rangle$	0
$\langle a, c \rangle$	0	$\langle b, c \rangle$	1	$\langle a, b, c \rangle$	0
$\langle a, d \rangle$	0	$\langle b, d \rangle$	1	$\langle a, b, d \rangle$	3
$\langle a, e \rangle$	0	$\langle b, e \rangle$	0	$\langle a, b, e \rangle$	2

$$S_x = w_a \cdot \sup(\langle a, x \rangle) + w_b \cdot \sup(\langle b, x \rangle) + (w_a + w_b) \cdot \sup(\langle a, b, x \rangle) \quad (5)$$

ここで、 $\sup(\langle l, m \rangle)$ は、行動パターン $\langle l, m \rangle$ のサポート値を表す。 w_k は、アイテム k を利用したことによる、ユーザの現在の価値判断への影響の重みを表す。時系列コンテキストとして現在の価値判断への影響の大きさは、現在に時間的に近いものほど大きいと考えられる。 w_k は、その影響の大きさを調整するためのパラメタである。

図1は、 w_k の定義を示したものである。横軸は、現在時刻を $t = 0$ としたときの相対的な時刻 t を表す。負の方向が過去を、正の方向が未来を示す。縦軸は、その相対的な時刻に依存する重み w の大きさを表したものである。 $t = 0$ のとき、 w は最大となり、現在時刻より離れるほど、 w は小さくなる。

w_k の定義の仕方はさまざまなものが考えられるが、図1には、例として3種類の定義を示しており、それぞれ次の意味を持つ。

- (a) 現在時刻より離れるほど、直線的に重みが減少する。
- (b) 現在時刻より離れるほど、指数関数的に重みが減少する。
- (c) ある閾値までは一定の重みを持つ。

どの定義の仕方が適切であるかは、実験的に求める必要がある。

図1(a)の重みの定義を用いると、各アイテムのスコアは次のように算出される。

$$\begin{aligned} S_a &= 0.4 \times 1 + 0.8 \times 1 = 1.2 \\ S_b &= 0.4 \times 2 = 0.8 \\ S_c &= (0.4 + 0.8) \times 3 = 3.6 \\ S_d &= 0 \\ S_e &= 0.4 \times 4 = 1.6 \end{aligned} \quad (6)$$

これより、アイテム c がユーザ A への推薦アイテムとして判定される。

5.4 ユーザコンテキストを考慮することの重要性

ここで、ユーザコンテキスト C'_A を次のように変える。

$$C'_A = (0.83, 0, 0.1, 1, 0, 0) \quad (7)$$

すると、ユーザ B の行動履歴データベースからデータ ID = {4, 5, 9, 10, 11, 12, 13, 14, 15}、ユーザ C の行動履歴データベースからデータ ID = {3, 4, 7, 8}、ユーザ D の行動履歴データベースからデータ ID = {8, 9, 10, 11, 12, 13}、をそれぞれ持つデータ集合が部分データベースとして扱われる。この部分データベースから同様に各行動パターンのサポート値を求めると、表5のようになる。

表5から、同様に各アイテムに対するスコアを算出すると、次

のようになる。 $S_a = 0, S_b = 1.2, S_c = 0.8, S_d = 4.4, S_e = 2.4$

これより、アイテム d, e が推薦アイテムとして判定される。以上から、同じ行動履歴データベースを対象としていたとしても、ユーザコンテキストが変わることにより、抽出される頻出行動パターンも変わり、結果、推薦されるアイテムも変わることが確認できる。

6. まとめと今後の課題

本稿では、過去/未来にユーザが取った/取りうる行動を時系列コンテキストとして扱い、この時系列コンテキストを考慮した情報推薦方式を提案した。特に、ユーザの時系列的な行動パターンがその時々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、その上で、過去/未来にユーザが取った/取りうる行動がユーザの現在の価値判断に及ぼす影響を考慮した情報推薦方式を提案した。

今後は、本提案方式を実装し、被験者を用いた評価実験を行う。評価実験を通じ、行動パターン抽出手法、重み付けの妥当性、アイテムに対するスコア算出方法の妥当性、について検証し、提案方式の有効性について考察を行う。

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)、および NICT 委託研究「電気通信サービスにおける情報信頼性検証技術に関する研究開発」による。ここに記して謝意を表します。

文献

- [1] 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-嗜好抽出技術の最新線-)”, 情報処理学会誌, Vol.48, No.9, pp.957-965, 2007.
- [2] G. Chen and D. Kotz: “A Survey of Context-Aware Mobile Computing Research,” Technical Report TR2000-381, Dept. of Computer Science, Dartmouth College, Nov. 2000.
- [3] 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.
- [4] 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), ACM, New York, NY, pp.20-31, 2000.
- [5] M. van Setten, S. Pokraev, J. Koolwaaij: Context-Aware Recommendations in the Mobile Tourist Application COMPASS. In Nejdli, W. and De Bra, P. (Eds.), AH 2004, LNCS 3137, Springer-Verlag, 2004.
- [6] 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, Vol.2007, No.16(20070301), pp.77-82, 2007-DPS-130-(14), 2007.
- [7] 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.
- [8] 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.
- [9] R. Agrawal and R. Srikant: “Mining Sequential Patterns,” In Proc. of ICDE1995, pp.3-14, 1995.
- [10] A. D. Wilson and A. F. Bobick: “Parametric Hidden Markov Models for Gesture Recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.21, no.9, pp.884-900, September, 1999.
- [11] 北 研二 (著), 津田 和彦 (著), 舞々 正幹 (著): 『情報検索アルゴリズム』, 共立出版 (2002).