

C-05

ユーザコンテキストを利用したユーザ嗜好及び
アイテム利用パターン分析に基づく情報推薦システム
Recommender System Based on Analyzing User Preference and
Item usage Pattern by Using Users' Contexts

関 匠吾† 張 建偉† 中島 伸介†
Shogo SEKI Jianwei ZHANG Shinsuke NAKAJIMA

1. はじめに

インターネット上には、膨大な量の情報が溢れており、この膨大な量の中から特定の情報を見つけるのは安易なことではない。このような背景から、ユーザに適切な情報の提供を可能にする情報推薦技術が注目されている。最近では、この情報処理技術に対して、ユーザの状況（コンテキスト）を考慮することで、推薦精度の向上を目指した取り組みがなされている。従来の情報推薦に関する研究で扱われるコンテキストの多くは、時間や場所など単純なコンテキストを扱うものであるが、奥らの研究[1]では、24次元のコンテキストを考慮したユーザの履歴データに SVM を用いて嗜好学習データを作成し、利用することで成果を挙げている。しかし、ユーザ毎に嗜好学習を行う為、新規ユーザもしくは利用経験が浅いユーザに対してはコールドスタート問題が発生する。

そこで我々は、ユーザがどのようなコンテキストの際にどのようなアイテムを好むのかを学習するのではなく、アイテム側の特性として、どのようなコンテキストのユーザから支持を受けているのかということを示すデータを蓄積することで、新規ユーザに対してもその時のコンテキストに応じて適切なアイテムを推薦することが可能な、アイテム利用時のユーザコンテキストを考慮した情報推薦システム提案している[2]。ただし、我々の行った提案では新規ユーザに対して既存のアイテムを推薦することが出来るが、新規アイテムに対するコールドスタート問題が発生するという欠点がある。

そこで我々は、ユーザ嗜好分析及びアイテム利用パターン分析をハイブリッド的に融合することで、新規アイテムと新規ユーザに対するコールドスタート問題を解決することが可能な情報推薦システムの提案を行う。本報では構築を目指すハイブリッド型のコンテキストを考慮した情報推薦システムの基本方針を紹介すると共に、アイテム利用パターン分析に基づく情報推薦手法に関する評価を行ったので報告する。

2. 関連研究と関連技術

ユーザの嗜好学習に関する研究として、奥ら[3]は、ユーザコンテキストごとに変化する価値判断基準に基づいたランキング手法を提案している。このシステムでは、学習データからユーザコンテキストに依存する価値判断基準モデルを使用し、そのモデルに基づいてランキングを行っている。また、協調フィルタリングを利用したものとして、Amazon.com[4]が有名であるが、これらはユーザの嗜好を学習するものであるため新規ユーザに対してはコールドスタート問題が発生する。

コンテキストを評価した技術として、食べログ[5]がコンテキストの評価分布を掲載している。この技術は、コンテキスト毎に星の数を用いて掲載している。しかし、コンテキスト毎であるため「夜にデート」といったように複合的に見る事が出来ない。

このように、コンテキストを考慮した研究や嗜好学習により推薦精度が向上したものはあるが、新規ユーザには精度の高い推薦が行えず、またコンテキスト毎の評価も複合

† 京都産業大学 コンピュータ理工学部, Faculty of Computer Science and Engineering, Kyoto Sangyo University

的に見ることができない。したがって、我々が提案する手法は新規性が高いと言える。

3. 提案手法

3.1 システムの概要

本手法は、ユーザコンテキストを利用したユーザ嗜好パターン分析とアイテム利用パターン分析の2つの分析により新規アイテムと新規ユーザに対するコールドスタート問題を解決するシステムである。解決する方法として、(1)の式を用いる。

$$S_T = \alpha S_U + (1 - \alpha) S_I \quad (1)$$

S_U はユーザ嗜好パターン分析による推薦対象アイテムに対するスコアを、 S_I はアイテム利用パターン分析による推薦対象アイテムに対するスコアを表している。また、 α は S_U と S_I のそれぞれの学習データ数の比により決定される値である。例えば、ユーザ嗜好パターン分析のためのデータ数が、アイテム利用パターン分析のための学習データ数に比べて、十分多い場合には、ユーザ嗜好パターン分析による情報推薦の重みを大きくするという方針で、 α の値を大きくすることになる。そして、それぞれのスコア合計である S_T を計算することができ、これに基づいて推薦対象アイテムをランキングすることにより、適切なアイテムの推薦を目指すものである。

ユーザ嗜好パターン分析に基づくスコア S_U は、奥らが提案する手法[3]をベースとして算出することを検討している。 S_I については、我々が独自に提案する手法に基づいて算出する。詳細は 3.3 節にて説明する。また、採用するユーザコンテキストやアイテムパラメータについては 3.2 節で述べる。なお、本章では推薦対象アイテムを飲食店として説明する。

3.2 採用するパラメータ

本手法では、ユーザコンテキストとして表 1 のように採用した。括弧内の数字は次元数であり、全てで 23 次元に座標軸を持つことになる。また、アイテム選定の際に邪魔になるとと思われるコンテキストを省くことにより効率の良い選定が可能となった。

飲食店の特徴としては、ホットペッパーAPI[6]を使用し、ての実装を行っているため、API から取得できる飲食店の情報をもとに表 2 のパラメータを採用した。

表 1 ユーザコンテキスト

日時情報	月 (2)	1 月～12 月
	時刻 (1)	05:00～25:00
気象情報	天気 (4)	晴れ/曇り/雨/雪
ユーザ情報	休日 (2)	休日/平日/平日前の休日/ 休日前の平日
	予算 (1)	1,000 円～10,000 円
	年齢 (1)	10 歳～60 歳
	性別 (1)	男性/女性
	経度 (東経) (1)	122 度 56 分 01 秒～153 度 59 分 11 秒
	緯度 (北緯) (1)	20 度 25 分 31 秒～45 度 31 分 35 秒
誰と/ 場面 (8)	1 人/家族/恋人/友人・先輩・後輩/上司・部下/ ビジネス/観光/その他	
人数 (1)	1 人～10 人	

表 2 飲食店の特徴

最寄り駅から	徒歩 0 分～10 分
経度(東経)	122 度 56 分 01 秒～153 度 59 分 11 秒
緯度(北緯)	20 度 25 分 31 秒～45 度 31 分 35 秒
営業時間	00:00～23:59
予算	1,000 円～10,000 円
ジャンル	居酒屋/ダイニングバー/創作料理/和食/洋食/ イタリアン・フレンチ/中華/焼肉・韓国料理/ アジアン/各国料理/カラオケ・パーティ/ バー・カクテル/ラーメン/お好み焼き・鉄板焼き /カフェ・スイーツ/その他
特徴	個室/座敷/貸切/駐車場/飲み放題/食べ放題/ お子様連れ

3.3 アイテム利用パターン分析によるスコア算出

以前我々が提案したアイテム利用時のユーザコンテキストを考慮した情報推薦システムの提案[2]を参考に述べる。

3.3.1 データ蓄積

検索から推薦、そしてデータ蓄積までの流れを例に表した図1を用いて説明する。なお、ここでのユーザは既存ユーザ、新規ユーザのどちらでも構わないとする。

- ① ユーザコンテキスト (UC) を取得して検索する。
- ② 推薦システムがユーザコンテキストを考慮して推薦アイテム一覧を提示する。
- ③ その推薦結果に対してユーザはアイテムを選択する。
- ④ ユーザが選択したアイテムとユーザコンテキスト (Data 1) をデータベースへ蓄積する。

蓄積する項目は、表3のようにアイテムのIDと3.2節で説明した表1のユーザコンテキストを0~1の範囲に数値化したものである。

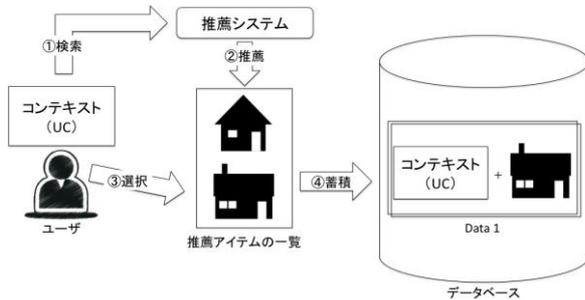


図1 データ蓄積

表3 蓄積データ例

	Data 1	Data 2
飲食店 ID	Item_A	Item_D
予算	0.3	0.5
年齢	0.366	0.533
...

3.3.2 スコア算出方法

スコア算出方法について説明するため、新規ユーザが利用した例である図2を用いる。

- ① ユーザコンテキスト (NUC) により検索する。
 - ② 蓄積データと NUC を 23 次元の空間座標に表し、k 近傍法により NUC を中心に近傍を探す。
 - ③ k 近傍法の k の個数により、飲食店 A のデータ□が多いことから飲食店 A を選定する。
 - ④ k 近傍法により選ばれた飲食店 A のデータからスコアの平均を取り、それを飲食店 A のスコア S_I とする。
- また、推薦アイテム一覧からユーザが選択したアイテムと NUC をデータベースへ蓄積する。

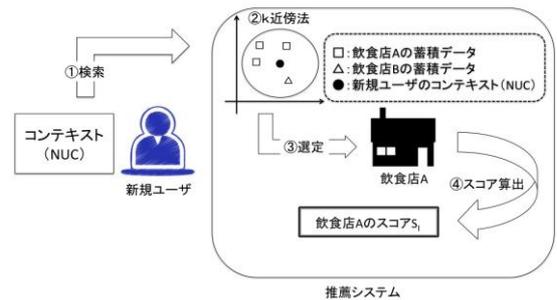


図2 S_I 算出までの流れ

4. 評価実験

提案する手法の有効性を評価するために、評価実験を行った。ユーザ嗜好パターン分析に関する有効性は、奥らの研究[3]により実証されているため今回は、アイテム利用パターン分析によるスコア算出について実験を行った。アイテムは京都市四条河原町付近の飲食店を使用する。また、ユーザコンテキストは表1より経緯と緯度を除いた21次元を使用した。

データ収集として、被験者9人に正解データとして、あるユーザコンテキストにおいて適切な店舗を選択してもらった。そのユーザコンテキストにより推薦されたアイテム一覧をもとに適合率 (P) と再現率 (R), F 値 (F) の平均スコアを算出した (表4)。

表4 適合率と再現率, F 値のスコア平均

P	R	F
0.519	0.556	0.536

実験結果より、ある程度の有用性は確認できたが、推薦

精度としては十分高いとはいえない。理由として、学習データに用いたデータ数が十分でなかったことと、評価実験で想定したコンテキストに偏りがあったことが挙げられる。したがって、今後はさらに詳細な評価実験を行った結果を踏まえてシステム実装に向けた検討を行う。

5. システムの運用に向けた検討

本章では、開発中システムの運用に向けた検討を行う。なお、推薦対象アイテムは飲食店として考える。

現在、システムの実装においてアイテム利用パターン分析については行えているが、ユーザ嗜好パターン分析についてはまだ行えておらず、スコア合計である S_T を算出することが出来ていない。また、データ数の比である α についてもどのように決定するかは検討段階である。これらについては今後、実験を通じてより精度の高いものになるよう実装して行く予定である。そして、評価実験にて行えなかった GPA を用いての経緯と緯度を含む評価も実装して行く予定である。

6. まとめ

本論文では、ユーザ嗜好パターン分析とアイテム利用パターン分析におけるスコアを算出することで、新規ユーザに推薦可能であり新規アイテムも推薦可能な、ユーザコンテキストを利用したユーザ嗜好パターン及びアイテム利用パターン分析に基づく情報推薦システムを提案した。

今後の課題として、推薦システムの実装及び評価実験に基づく手法の改良を行う。

謝 辞

本研究の一部は、文部科学省科学研究費助成事業（学術研究助成基金助成金）基盤研究（C）（課題番号：#23500140）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: 状況依存型ユーザ思考モデルに基づく Context-Aware 情報推薦システム, 情報処理学会, データベース, Vol. 48,

No. SIG(TOD 34), pp. 162-176 (June 2007).

- [2] 関匠吾, 張建偉, 中島伸介: アイテム利用時のユーザコンテキストを考慮した情報推薦システムの提案, DEIM Forum 2011, B1-1(2011)
- [3] 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, Vol. 2, No. 1(TOD 41), pp57-80 (Mar 2009).
- [4] 『Amazon.com』. <http://www.amazon.com/>.
- [5] 『食べログ』. <http://tabelog.com/>.
- [6] 『ホットペッパー Web サービス』. <http://api.hotpepper.jp/>.