

嗜好の遷移過程に着目した Web 推薦アルゴリズムの提案

A Web Recommendation Algorithm Based on Transition Models

古川 忠延*¹ 松尾 豊*² 内山 幸樹*³ 石塚 満*¹
 Tadanobu Furukawa Yutaka Matsuo Koki Uchiyama Mitsuru Ishizuka

*¹ 東京大学大学院情報理工学系研究科

Graduate School of Information Science and Technology, The University of Tokyo

*² 東京大学大学院工学系研究科 *³ 株式会社ホットリンク

Faculty of Engineering, The University of Tokyo

Hotto Link Inc.

We propose a recommendation algorithm based on transition patterns of users' adoption. This algorithm uses dataset-independent features, and aims to give a general-purpose method.

1. はじめに

近年のウェブコンテンツは、ウェブ掲示板やブログの他、mixi *¹に代表されるソーシャル・ネットワーク・サービスや、YouTube *²のような動画共有サービス、はてなブックマーク *³や del.icio.us *⁴などのソーシャル・ブックマークなどの流行に見られるように、増加・多様化の一途を辿っている。ウェブを閲覧するユーザにとって、魅力的なコンテンツの増加は好ましい事態であるが、反面、情報の増加により有用なコンテンツへのアクセスが困難になることも危惧される。こうした情報の氾濫への解決策の一つとして、これまで多くのコンテンツ推薦手法が提案されてきた。多くは内容の類似性や統計的な行動傾向、世間の流行・評判などに基づいて情報を推薦しようというものであり、特に EC サイトでは推薦サービスを積極的に取り入れられてきた。しかし、対象とするコンテンツやユーザによっては、的確な推薦を行うことは困難であり、様々なアルゴリズムが提案されている [Adomavicius 05]。

本稿では、様々な種類のウェブコンテンツに対して、汎用的に利用できる推薦アルゴリズムを提案する。コンテンツに独特な属性を用いず、「ユーザがある時間にアイテムを採用する (= 閲覧する、購入する、ブックマークする)」という基本的なデータのみから推薦を行うものである。

以下、まず 2 章において既存の推薦手法を紹介する。3 章にて提案手法を説明、4 章で評価実験を行う。最後に、5 章にて本稿をまとめる。

2. 既存の推薦手法

推薦システムにおいて用いられる主な手法は以下の二つである。

- 協調フィルタリング: 採用履歴からユーザ間の類似度を測り、類似ユーザの採用アイテムを推薦する手法。履歴が少ない場合には類似度の信頼性が下がる問題点がある。

連絡先: 古川 忠延, 東京大学大学院 情報理工学系研究科 創造情報学専攻, 〒101-0021 東京都千代田区外神田 1-18-13 秋葉原ダイビル 13F, furukawa@mi.ci.i.u-tokyo.ac.jp

*¹ <http://mixi.jp/>

*² <http://www.youtube.com/>

*³ <http://b.hatena.ne.jp/>

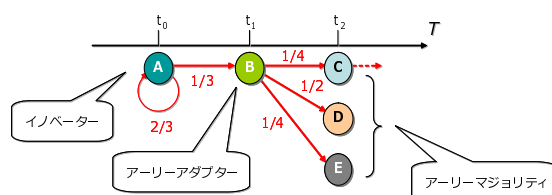
*⁴ <http://del.icio.us/>

- 内容に基づくフィルタリング: ユーザが採用したアイテムと内容の類似したアイテムを推薦する手法。アイテムの情報が必要な他、採用したことの無い、意外性のあるアイテムを推薦することができない。

本稿で提案する手法は、アイテムの採用過程から平均的なアイテム採用の伝播傾向に着目するものであり、上記の協調フィルタリングに近い方法である。採用の時間を考慮し、ユーザ・アイテムの伝播力を計算することで予測精度の向上を目指す。

3. 提案手法

本稿では、[Song 07] の手法を改良する形のアルゴリズムを提案する。[Song 07] では、あるアイテムをユーザが採用する過程を連続時間マルコフ連鎖 (CTMC) と捉え、さらにイノベーション普及モデル [Rogers 95] に基づいてユーザ間でのアイテム伝播率を計算することで、アイテムの普及をモデル化している。ユーザ間でのアイテム採用の伝播傾向を利用することで、時間 t においてユーザがアイテム r を採用しているとき、時間 $t + \Delta t$ においてアイテム r を採用するであろうユーザを予測することができ、推薦システムに利用することが可能となる。



- 推薦
 - 次に誰がアイテムに興味を持ちそうか
→ 興味を持ちそうな人に対して推薦
 - アイテムは人気が出そうな状況かどうか
→ 人気になりそうなら推薦

図 1: アイテム採用伝播のマルコフ連鎖による表現

このモデル化を数式で表現すると、以下のようになる。まず、CTMC における遷移について、ユーザ i からユーザ j へのアイテムの遷移確率を指数関数で仮定し、遷移にかかった時間 t から伝播率 q_{ij} を計算する (図 2)。同一ユーザ間の遷移

率 q_{ii} は、最速での伝播に要した時間の逆数で定義する．そして、得られた遷移率行列 Q より、遷移確率行列 $P(t)$ を計算、これを積分することによって、ある時間における遷移状態を予測することができる (式 (2)) ．

$$Q = \begin{pmatrix} q_{0,0} & q_{0,1} & \dots \\ q_{1,0} & q_{1,1} & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} = \begin{pmatrix} -q_0 & q_{0,1} & \dots \\ q_{1,0} & -q_1 & \dots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

$E(X_i) = \frac{1}{q_i} = T_i$

$$P_{ij} = \sum_c q_i \exp(-q_i t_{ij}(c))$$

$$q_{ij} = q_i P_{ij} \quad (i \neq j)$$

図 2: 遷移率行列の計算

$$P'(t) = P(t)Q, \quad P(0) = I \tag{1}$$

$$L(\Delta t) = \int_0^{\Delta t} P(t)dt \tag{2}$$

この手法では、全てのアイテムについて、データの平均をとって伝播モデルを構築している．そのため、アイテムごとの特徴を考慮するためには、あらかじめアイテムの属性を利用したクラスタリングが必要である．そこで本稿では、アイテム間の伝播の推移も同様に CTMC によるモデル化を行う (図 3) ．これにより、「どのアイテムの採用後は、どのアイテムを採用しやすいか」という予測をすることができる．

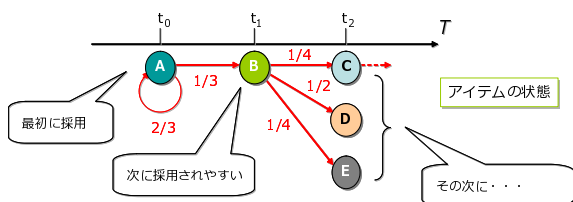


図 3: 採用アイテム遷移のマルコフ連鎖による表現

4. 実験

提案手法の評価実験として、はてなブックマークのデータを用いて実験を行った．データの期間は 2003 年 4 月から 10 月まで、対象ユーザは 1201 人、対象アイテムは 2114 の URL である．はてなブックマークはソーシャル・ブックマークの一つであり、ユーザが興味を持ったウェブサイトを、オンライン上でブックマークするサービスである．実験では、期間の半分 (6 月まで) のデータから、アイテム間・ユーザ間の二種類の遷移確率を計算し、残りの期間 (テスト期間) でブックマークするであろうサイトの予測を行う．また、本手法は採用の遷移パターンが予測しやすいアイテム (シリーズ物や同一カテゴリーのアイテムなど) に有効であろうことを考慮し、拡張として対象アイテムの限定や、特定のアイテム間の遷移への重み付けをしての実験も行った．

結果を表 1 に示す．精度は、テスト期間に採用する可能性が高いと判定されたアイテム上位 10 のうち、実際にユーザが採用していたアイテム数の割合である．拡張をしていない場合、データ数が少なく、特にアイテム間の遷移データは疎であり、

表 1: 予測精度

手法	ユーザ遷移	アイテム遷移	組合せ
拡張なし	5.2 %	2.0 %	5.2%
アイテム限定	7.1 %	8.4 %	8.9%
タグ使用	5.7 %	5.5 %	6.3%

また、アイテムは短い期間のうちにブックマークされることが多かったため、アイテム遷移による予測精度は非常に低いものとなった．両者を組み合わせた (遷移確率を足し合わせて予測を行った) 場合も、ユーザ遷移による予測と同等の結果が得られるにとどまっている．

一方で、拡張を加えた場合には精度の改善が見られている．スポーツ・政治など、特定の話題に関するアイテムのみを使用することで評価を行った場合、アイテム間遷移による予測がユーザ間遷移による精度を上回った．また、類似したタグが付与されたアイテム間の遷移に対して、その遷移パターンに重み (一律 2 倍とした) を加えた場合にも、両遷移を組み合わせた場合はユーザ間遷移のみによる精度を上回り、ユーザ間遷移だけでは予測できないアイテムの予測に成功している．

5. まとめ

本稿ではユーザ間のマルコフ連鎖に基づいた既存の推薦手法を基に、アイテム間のマルコフ性を考慮することで、多くのデータに対応する推薦アルゴリズムを提案した．遷移パターンがスパースなデータにおいては効果を得られなかったが、アイテムの内容情報を考慮して対象を限定することで精度の改善が見られ、データ次第では有効な手法となり得ることがわかった．今後はアルゴリズムの汎用性を目指し、アイテムの内容情報を用いずに改良していく方法を模索していきたい．

参考文献

[Adomavicius 05] Adomavicius, G. and Tuzhilin, A.: Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions, *Knowledge and Data Engineering, IEEE Transactions on*, Vol. 17, No. 6, pp. 734-749 (2005)

[Rogers 95] Rogers, E. M.: *Diffusion of Innovations*, The Free Press (1995)

[Song 07] Song, X., Chi, Y., Hino, K., and Tseng, B. L.: Information Flow Modeling based on Diffusion Rate for Prediction and Ranking, in *Proc. WWW 2007* (2007)