

ユーザ革新性の推定手法に関する評価 An Evaluation for Estimating Users' Innovativeness

中村 美穂^{*1}
Miho NAKAMURA

市川 裕介^{*1}
Yusuke ICHIKAWA

後藤 真一郎^{*1}
Shinichiro GOTO

黒川 裕彦^{*1}
Hirohiko KUROKAWA

中川 哲也^{*1}
Tetsuya NAKAGAWA

^{*1} 日本電信電話株式会社 NTT 情報流通プラットフォーム研究所
NTT Information Sharing Platform Laboratories, NTT Corporation

A recommender system has been paid attention because of information explosion. We have been studied a recommendation method considering users' Innovativeness. Innovativeness is one of the users' preferences and means users' acceptance for new items. We classify users some groups which is defined by Rogers as "diffusion of innovations theory". In this paper, we propose a classifying method users' Innovativeness using Web action data and an assumption about the relationships between users' Innovativeness and Web action data. The purpose of this study is to evaluate an assumption using a questionnaire. This study describes that our assumption might be proved.

1. 序論

Web 上で流通する情報の爆発的な増加により、推薦システムに注目が集まっている[市川 08]. ユーザの情報選択を推薦システムでサポートするには、ユーザが実際に情報選択時に重視するポイントを考慮して、情報を推薦することが必要である.[Yahoo!リサーチ 05]によると、ユーザはコンテンツの新規性や話題性を重視して選択すると報告されている。

本研究では、ユーザの新規性や話題性に対する嗜好を考慮した推薦を実現するため、ユーザが持つ「革新性」に着目し、推薦方式の提案を行ってきた[中村 08a]. 革新性とは、新コンテンツに対する受容性を表す特性であり、イノベーター理論として広く知られている[Rogers 62]. イノベーター理論ではユーザの革新性に応じて、5 つのユーザグループ (Innovator, Early Adopter, Early Majority, Late Majority, Laggards) を定義している。中でも、Innovator と Early Adopter は新規性を重視するユーザ、Early Majority と Late Majority は話題性を重視するユーザに該当する。従って、革新性を取り入れることにより、新規性や話題性に対する嗜好を考慮することが可能になる。

革新性に応じた推薦を実現する為には、まずユーザの革新性を知る必要がある。そこで本稿では、ユーザの革新性を推定する方式と仮説の提案、及びその検証結果について報告を行う。

2. 関連技術

既存技術は、主に以下に示す方法で革新性を推定している。

2.1 アンケートによる推定

マーケティング分野では、主にアンケートデータを用いてユーザの革新性を推定してきた。本手法を推薦システムに適用した場合、推薦を受けようとする全てのユーザがアンケートに回答する必要があり、ユーザに嗜好の入力負担がかかる点が課題である(課題 1)。

2.2 ログデータによる推定

(A)コンテンツを採用したタイミングに着目して推定する手法[石川 07][毛受 08]

革新性に応じて、ユーザがコンテンツを採用するタイミングが異なることに着目した手法である。本手法は、商品の発売後すぐに採用する傾向を持つ Innovator の抽出には有効であるが、Innovator 以外のユーザを抽出するには、長期間のログデータが必要になる(課題 2)。理由は下記の通りである。

Innovator から Laggards のユーザを抽出するには、全グループの採用タイミングに関するルールを予め知っておく必要がある。例えば、ルールとは下記のようなものを指す。

(ルール 1)コンテンツの発売開始日から X_1 日目までに採用したユーザは Innovator と判断する。

(ルール 2)コンテンツの発売開始後 X_1 日目から X_2 日目までに採用したユーザは Early Adopter と判断する。

イノベーター理論によると、ユーザの革新性は全採用者数の中で採用が早い順に Innovator から Laggard までを定める。従って、 X_1 や X_2 は「全採用者数」に対する「 X_1 や X_2 までの採用者数」の割合によって決まる。「全採用者数」と「各時点での採用者数」を知るには、コンテンツの発売から販売終了までの長期間のログデータを分析する必要がある。一方、推薦システムは少ないログで適切な情報を推薦できることが望ましい。本手法は、長期間のログの蓄積がないとユーザの革新性を推定することが困難である点が課題である。

(B)ユーザの採用順序に着目して推定する手法[Rusmevichientong 04]

同じコンテンツを採用したユーザの相対的な順番に着目して、ユーザの革新性を推定する手法である。本手法は、Innovator 以外のユーザを抽出するには、長期間のログを蓄積しないと革新性の推定が困難である(課題 2)。理由は下記の通りである。

本手法は、コンテンツの全採用者数のうち、何番目にユーザがコンテンツを採用したかを導出してユーザの革新性を推定する。従って、(A)の手法と同様、採用者数を知る為、コンテンツの発売から普及終了までの長期間のログデータが必要となる。

2.3 関連技術の課題

以上より、関連技術には以下の課題が存在する。

連絡先: 中村美穂, NTT 情報流通プラットフォーム研究所,
東京都武蔵野市緑町 3-9-11, Tel:0422-59-3119,
Fax:0422-59-5657, Email:nakamura.miho@lab.ntt.co.jp

- 課題 1: ユーザに嗜好の入力負担がかかる。
- 課題 2: 長期間のログデータが必要になる。

3. 提案手法

3.1 仮説

革新性, ユーザのパーソナリティ, Web 行動は図 1 に示す関係をもつと考え, 下記の仮説を提案する。

[仮説]各種 Web 行動と革新性は相関を持ち, Web 行動を観測することによりユーザの革新性を推定できる。

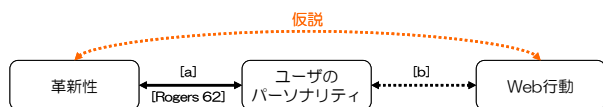


図 1. 革新性, パーソナリティ, Web 行動の関係

図 1 の[a]には, 具体的には下記のパーソナリティが革新性によって異なることがわかっている[Rogers 62].

- P1: 不確定性, 危険, 変化に対する態度
- P2: 情報の収集に対する態度, 知識量
- P3: オピニオンリーダーシップ
- P4: 社会的な地位 (収入など)

本研究では, パーソナリティと対応する Web 行動として表 1 の Web 行動を挙げ, ユーザのパーソナリティによって a から k の Web 行動が異なると仮定した(図 1 の[b]).

表 1. パーソナリティと対応する Web 行動

パーソナリティ	Web 行動
P1	a.サイトのオススメ商品を重点的にチェックする b.商品のレビュー(評価)をチェックする c.商品の評価が高い順にソートしてチェックする d.売れ筋商品のランキングページをチェックする e.新発売の商品をチェックする
P2	f.EC サイト内の多くのページを閲覧する g.EC サイトのトップページのみを閲覧する h.詳しい商品解説のあるページをチェックする i.サイト内の検索機能を使って商品を探す
P3	j.購入した商品のレビューや評価を書き込む
P4	k.商品の価格が高い順にソートしてチェックする

3.2 概要

ある 1 サイト内での Web 行動から革新性を推定する手法を提案する。提案手法の概要を図 2 に示す。まず, アクセスログデータを用いて Web 行動データを算出する。Web 行動データとは, 表 1 に示した各 Web 行動の行動頻度を記したデータである。次に, Web行動データに対して推定モデルを適用し, ユーザの革新性を推定する。なお推定モデルは, 革新性が別手段によって判明した一部のユーザのデータを教師データとして使用する, 半教師つき学習によって生成する。

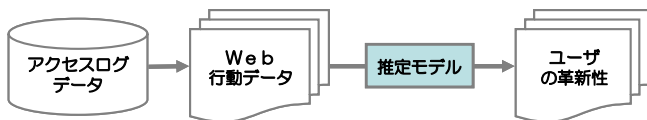


図 2. 提案手法の概要

3.3 効果

提案手法は, 既存技術の課題に対して下記のように解決する。

課題 1: アクセスログデータを用いてユーザの嗜好を自動推定する。例えば, 半教師つき学習であっても全員にアンケートを

実施する必要がないため, ユーザへの負担を軽減する効果が期待される。

課題 2: 提案手法に必要な Web 行動データは, 数セッション~数日で取得可能である見込みが高く, コンテンツの普及が収束するほど長期間(数ヶ月~数年)のデータは必要ない。

4. 検証

Web アンケートデータを用いて, 仮説検証を行った。

4.1 検証方法の概要

交差検証法を用いて推定モデルの生成, 及び推定精度の算出を実施した。本検証では, 推定精度より仮説が成立するか否かを確認した。Web 行動データからユーザの革新性を高い精度で推定することが可能な場合には, 仮説が成立する可能性が高いと判断した。

検証方法の概要を図 3 に示す。まず, 一部のユーザの Web 行動データと他の手段で推定した革新性データを用いて推定モデルを生成した(1)。なお, 推定モデルの生成は本検証では線形判別分析を採用した。次に, 推定モデルを用いてモデル生成に用いていないユーザの革新性データを推定し(2), 他の手段で推定した革新性データと比較することにより推定精度を算出した(3)。

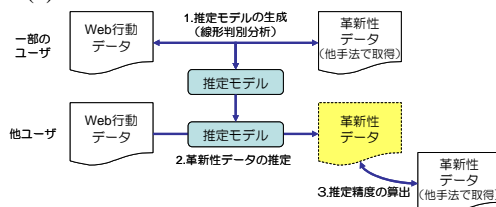


図 3. 検証方法の概要

本検証では推定モデルを生成する為に, 教師データとなるユーザの革新性データが必要である。そこで, マーケティング分野で利用されている既存の手法を用いて革新性データを生成した。既存の手法とは, 株式会社構造計画研究所が作成した IOF 尺度で, ユーザにイノベーター, オピニオンリーダー, フォロアーの各特性に関する計 6 つの質問を行い, 各質問に対する回答結果の組合せからユーザを 4 グループ (Innovator, Opinion Leader^{*1}, Follower^{*2}, Others^{*3}) に分類する手法である。具体的な質問は 4.1 の Q1 の通りである。

^{*1}Opinion Leader は, イノベーター理論の Early Adopter に該当する。

^{*2}Follower にはイノベーター理論で定義された Early Majority, Late Majority, Laggards に該当する。

^{*3}Others は, 革新性に基づく手法では分類できないユーザに該当する。

4.2 アンケートデータ

2007 年 9 月にアンケートを実施し, 化粧品を EC で購入した経験を持つユーザ 49 名(男性 5 名, 女性 44 名)から下記の Q1, Q2 に対する回答を収集した。本検証では, 化粧品を購入した経験をもつユーザの多くが革新性^{*}を持つこと[中村 08b]を考慮して, 検証対象を化粧品と定めた。(※[中村 08b]では, 革新性を採用時期特性と表現しており, 同じ嗜好特性を表す。)

Q1. i ~vi の各項目に対し, 次の選択肢から回答してください。
{非常にあてはまる, ややあてはまる, あまりあてはまらない, 全く当てはまらない}

- i. 人より新しいものを買う
- ii. 新製品は 1 度は買ってみる
- iii. 製品の特徴を他人に知らせる
- iv. 宣伝で見る製品を選ぶ

- v.気に入っている商品を他人に薦める
 - vi.周囲が使用している製品を選ぶ
- Q2. 化粧品の EC サイトにアクセスする場合、表 1 に示した a ~k の行動をどのくらいの頻度で行いますか？
次の選択肢から回答してください。
{ほぼ毎回, 2 回に 1 回, 3~4 回に 1 回, 5~6 回に 1 回, それ以下, しない}

4.3 手順

- ①Q1 の回答データを用いて、ユーザの革新性データを生成した。
- ②Q2 の回答データを用いて、ユーザの Web 行動データを生成した。定量的な分析を行う為に、Q2 の各回答に対して以下のように 1~6 の数値を割り当てた。
ほぼ毎回→1, 2 回に 1 回→2, 3~4 回に 1 回→3, 5~6 回に 1 回→4, それ以下→5, しない→6
- ③統計ソフト SPSS 16.0J を用いて線形判別分析を実施し、推定モデルを算出した。さらに、交差検証法を用いて推定精度を算出した。

4.4 結果

4.3 の手順①(既存手法)で推定したユーザの革新性の分布を図 4 に示す。今回の検証ユーザは Innovator が少なく、Opinion Leader や Follower はそれぞれ 10 数名のユーザが存在することを確認した。本検証では、Others を除く 40 名を検証対象とした。

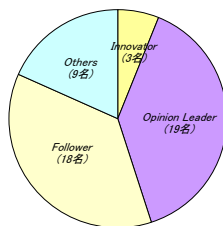


図 4. 既存手法によるユーザの革新性の分布

全ユーザのデータを用いて推定モデルを生成し、適用した場合のクローズドな推定精度は 80.0%であった。また、交差検証法で算出した推定精度は 55.0%であった。表 2 は、交差検証法で推定した結果と既存手法で推定した結果の一致状況を示している。表 2 で色付けを行ったセルが、既存手法と推定手法で同じ結果を示した箇所で、40 名中 22 名 (Innovator 1 名, Opinion Leader 10 名, Follower 11 名) のユーザが既存手法と同じ推定結果を示した。

表 2. 交差検証法による推定結果

		提案手法			合計 [名]
		Innovator	Opinion Leader	Follower	
既存 手法	Innovator	1	2	0	3
	Opinion Leader	2	10	7	19
	Follower	1	6	11	18

4.5 考察

既存手法による推定結果に対して、検証に用いた手法は約半分の推定精度で Web 行動からユーザの革新性を推定することが可能であることを確認した。検証対象が 3 グループであることからランダムに推定した場合、推定精度は 33%前後と見込まれる。ランダムな場合の推定精度(33.0%)と検証に用いた手法の推定精度(55.0%)を比較すると、今回の推定精度は高い値を示しており、仮説が成立する見通しを得た。ただし、本検証は

Innovator のサンプル数が少ないため、推定精度の有意性には注意が必要である。

今回の検証では簡便な手法である線形判別分析を実施して推定モデルを生成しており、今回のデータに適切な手法であったか考察が必要である。

また、推定モデルの生成に用いたデータに対する推定精度と、交差検証法で算出した推定精度の間に 25%の差が存在していた。これは、推定モデルが過剰適合を起こしている可能性を示唆している。過剰適合が生じる原因として、以下の理由が考えられる。

- (1)推定モデルを説明する Web 行動の数が多い。
- (2)推定モデルの生成に用いたデータに偏りがあった。

今後過剰適合を抑える為に、下記のような取り組みが必要である。

- (1)Web 行動間の相関を確認し、相関の高い Web 行動はモデルの推定からは外す。
- (2)多くのサンプルデータを集めて、生成に用いるデータの偏りを減らす。

今後は、上記で述べたことを踏まえ、実際のアクセスログデータを用いた Web 行動データによる推定方式の検証と、推定したユーザの革新性を推薦方式に適用した際の有効性について検証を実施する予定である。

5. 結論

本稿では、推薦方式に必要な革新性の推定方式を提案し、アンケートデータを用いて Web 行動の集計データからユーザの革新性を推定できる見通しを得た。今後は、実際のアクセスログデータを用いて提案手法の検証を行う予定である。

参考文献

[石川 07] 石川雅之, 森田武史, 和泉憲明, 山口高平: 共有コンテンツのアクセス履歴分析に基づく情報推薦エンジン, 第 21 回人工知能学会全国大会, 3G8-2, 2007.

[市川 08] 市川裕介, 中村美穂, 畑恵介, 中川哲也: サイト横断でユーザ個々の嗜好に合ったサービス提供を実現するパーソナルサービス基盤技術, NTT 技術ジャーナル, Vol.20 No.1, pp.67-70, 2008.

[中村 08a] 中村美穂, 市川裕介, 後藤真一郎, 黒川裕彦, 中川哲也: ユーザの採用時期特性を考慮した推薦方式の研究, 第 19 回データ工学ワークショップ(DEWS2008), B7-5, 2008.

[中村 08b] 中村美穂, 市川裕介, 後藤真一郎, 黒川裕彦, 中川哲也: カテゴリ横断型の推薦を実現するユーザプロファイルの提案, 電子情報通信学会 2008 年総合大会, D-8-23, 2008.

[毛受 08] 毛受崇, 吉川正俊: ブックマークの時系列情報を利用したソーシャルブックマークにおける注目度予測, 第 19 回データ工学ワークショップ(DEWS2008), B9-5, 2008.

[Rogers 62] E.M.Rogers: Diffusion of Innovations, The Free Press, NewYork, 1962.

[Rusmevichientong 04] P.Rusmevichientong et al.: Identifying Early Buyers from Purchase Data, In Proc. of the ACM SIGKDD, pp.671-677, 2004.

[Yahoo!リサーチ 05] ヤフー株式会社. “インターネットショッピング市場現在と今後の展望～2005 要約レポート,” <http://i.yimg.jp/images/research/pdf/reportssummary20050726.pdf>, Yahoo!リサーチ, 2005-07-26, (参照 2008-04-07).