

Webサイトにおける項目反応理論を用いたおすすめコンテンツの提示 Suggestion for Recommended Contents on Website with Item Response Theory

橋間 智博†

Tomohiro Hashima

1. はじめに

近年、インターネット利用者の裾野の広がりは激しく、今後もさらに拡大していくことは、容易に予想ができる。ライトユーザーの人口が増えることで、サイトの利便性向上の重要性はますます高まり、物販サイト等を含めたサイト運営側が「訪問者に提示したいページを見てもらっているのか?」という不安が増している。

本研究では、訪問者の嗜好を調査し、Webサイトにおけるナビゲーションの利便性向上を目的とし、研究を行った。項目反応理論を用いることでWebサイトのアクセス分析を行い、各ページの特性を求める。そして、リアルタイムで訪問者の閲覧履歴から各ページとの関係性を求め、訪問者が求めていると予想されるページを提示することを目標としている。

2. 項目反応理論によるアクセス分析

本章では、項目反応理論 [1] のアクセス分析への適応方法について述べる。

2.1 分析対象 Web サイト

対象とした Web サイトは兵庫県立大学経済学部 Web サイト (以下、対象サイトと呼ぶ) である。対象サイトは CMS(Content Management System) である XOOPS を用いて構築されており、本研究では XOOPS モジュールを作成し、実験を行った。

アクセス分析には文献 [2] で筆者が作成した、アクセスログ分析ツールを用いて集計を行った。

2.2 項目反応理論によるアクセス分析の適用

まず初めに、対象サイトのアクセスログを、文献 [3] に従い項目反応理論を用いたアクセス分析を行った。表 1 にアクセスログ収集時の条件等を示す。

表 1: アクセスログの条件

期間	2007年1月15日から2月13日
アクセス数	5,470件
ページ数	78ページ

各ページを項目とし、訪問者のページ閲覧の有無から反応パターンベクトル U_{ij} を生成した。そして、ベイズ推定法を用いて各ページの項目識別力と項目困難度を求めた。さらに、検索エンジン経由での訪問による検索キーワードも項目として採用した。

2.3 マルコフモデルとの比較

項目反応理論を用いたアクセス分析 (以下、項目反応モデルと呼ぶ) による、おすすめコンテンツの提示方法と比較するために、マルコフ過程を用いた提示方法 (以下、マルコフモデルと呼ぶ) を同時に実施した。表 1 のアクセスログを用いて、サイト内のページからページへの遷移確率を求めた。

3. 実装

対象のサイトは、XOOPS を用いて構築されており、本研究による機能と検証用のデータ収集は、XOOPS のモジュール (以下、モジュールと呼ぶ) として実装した。

3.1 コンテンツ適合度

訪問者によってページが表示される毎にモジュールが呼び出される。モジュールが呼び出されると、おすすめコンテンツを示す 5 つのリンクが表示される。訪問者へのおすすめコンテンツの表示は、コンテンツ適合度によって示される。コンテンツ適合度はページが表示される毎に、項目反応モデルかマルコフモデルのどちらかで算出される。どちらのモデルで算出されるかは、それぞれ二分の一の確率で決定する。

項目反応モデルによって算出する場合、まず、訪問者が表示しているページのアドレスを閲覧履歴に追加する。さらに、検索キーワードを持つ訪問者の場合には、検索キーワードをアドレスに置き換えて閲覧履歴に追加する。そして、2.2 で算出した項目識別力と項目困難度を用いて、閲覧履歴から訪問者の特性値 θ を推定する。訪問者の特性値 θ を用いて、2.2 で算出した項目識別力と項目困難度からすべてのページの項目情報量を求め、それぞれのページの訪問者に対するコンテンツ適合度とした。

マルコフモデルによって算出する場合、2.3 であらかじめ求めた遷移確率を用いて、表示されたページからサイト内の各ページへの遷移確率を求め、これをコンテンツ適合度とした。

3.2 類似コンテンツへのリンクの表示

項目反応モデルかマルコフモデルによって求められたコンテンツ適合度上位 5 位のリンク (以下、リンクと呼ぶ) を、モジュールの表示ブロックに表示させる。

図 1 に、モジュールを表示させた Web サイトのスクリーンショットを示す。それぞれのリンクには、ページタイトル (ページの属するジャンルとコンテンツのタイトル) とコンテンツ適合度を棒グラフで表示した。棒グラフに関する説明は特に表示していない。モジュールはページ左側の訪問者別メニューや、バナーと縦並びで表示した。リンクのフォントは、ページ本文のフォントよりも 2 ポイント小さくなっており、訪問者別メニューとくらべると、かなり小さくしている。

3.3 検証用データの収集

モジュールが表示される毎に、項目反応モデルとマルコフモデルのどちらでコンテンツ適合度を算出したのか、回数を計測した。また、リンクが実際にどの程度クリックされたかを後のアクセス分析で計測するために、アンカータグの href 要素にクエリ文字列として、項目反応モデルによる表示の場合「fr=ia」、マルコフモデルによる表示の場合「fr=ma」を付加した。

†兵庫県立大学経済経営研究所



図 1: リンクの表示

4. 運用後のアクセス分析と評価

本章では, 3. で作成したモジュールを, 対象サイトに設置し, 実際に1ヶ月間運用した後のアクセス分析と, モジュールで表示したリンクのクリック率を集計し, 項目反応理論によるおすすめコンテンツの提示についての評価を述べる.

4.1 モジュール追加後のアクセス分析

3. で作成したモジュールを, サイト上で実際に1ヶ月間運用した. ログを収集した期間は, 2007年3月6日から4月5日の1ヶ月間で, アクセス数は5,558件であった. また, モジュールが表示されるのは, 2ページ目からであることから, 閲覧ページ数が1ページのみの訪問者は, 今回の評価には無関係であるので除外する. 除外した結果, 有効なアクセス数は3,560件, ページの総表示回数は26,462回であった.

4.2 項目反応モデルとマルコフモデルへの反応

項目反応モデルによって提示されたリンクのクリック回数は, 表示回数11,716回に対して, 489回(4.17%), マルコフモデルによって提示されたリンクのクリック回数は, 表示回数11,154回に対して, 336回(3.01%)であった. 両モデルによる表示回数の合計が, 4.1のページの総表示回数よりも少ないが, これは, 項目反応モデルの場合に, 項目識別力が低く, 項目母数が求められなかったアドレスへのアクセスによるものである. また, マルコフモデルの場合, 2.3にて遷移確率を求めたが, その際にアクセス数が非常に少なく, 遷移確率を求めることができなかったアドレスへのアクセスによるものである.

4.3 項目反応モデルの評価

対象サイトのように, コンテンツ量が少なく, サイト構造が浅いツリー構造になっており, トップページからメニューを介して, すべてのコンテンツにアクセス可能であるサイトにおいて, ショートカットの様な役割を持つモジュールの存在意義は小さい. また, 入試情報などのグループについては, 一通り順番に見てしまうという訪問者も多い.

両モデルによって表示されたリンクのクリック数については, 数字としては少ないように思えるが, 筆者の感想としては, マルコフモデルでのクリック率3%は予想通りであった. 物販サイトにおける, ページの両脇などに表示される「おすすめ商品」へのリンクや, 「他の人は一緒にこんな物を買っています」などのクリック率と比べてみても, 遜色のない結果だと考える.

また, 項目反応モデルによるクリック率は, 予想よりも少し低かった. しかし, 項目反応モデルとマルコフモデルのクリック率の比は約138.6%と, コンテンツ適合度を項目反応モデルで算出した場合が, マルコフモデルに比べクリック率が高くなっている. したがって, マルコフモデルに比べて, 項目反応モデルの方がコンテンツ適合度をより適切に算出でき, おすすめコンテンツを求めるためには, 適していると考えられる.

5. まとめと今後の課題

本研究では, 項目反応理論を用いて, アクセスログからWebサイト内における各ページの項目母数を求め, 訪問者の特性値 θ から項目情報量を求め, 訪問者が求めているであろうページを提示することを目的とし, 研究を行った.

本研究の結果から, 項目反応モデルによるおすすめコンテンツの提示は, マルコフモデルによる提示よりも訪問者の反応が高かった. したがって, 項目反応理論を用いて, 訪問者が求めるコンテンツを推定することが可能だと思われる.

今後の課題としては, 以下の点があげられる.

リンク表示の改善

本研究での表示方法では, リンクが他のメニューと独立性が高く, 浮いてしまっていることから, 他のメニューに溶け込むような表示方法を考える必要がある.

評価データ収集方法の改善

本研究では, ページが表示される度に項目反応モデルかマルコフモデルを使うかを決めていたが, 訪問者毎に決めるなどの工夫が必要である.

多値データでの分析

本研究では, 各ページについて訪問者の閲覧の有無だけの, 2値データモデルとして扱った. しかし, ページ数が多く, メニューが多階層になっている場合などは, 同じページを繰り返し閲覧することになる. こういった場合, 多値項目反応モデルを用いるなどの工夫を用いることで, ページの特性を求めることが容易になるのではないかと考える.

謝辞 本研究を進めるにあたり, ご指導頂いた兵庫県立大学経済学部秋吉一郎教授に心より感謝いたします.

参考文献

- [1] 豊田秀樹, "項目反応理論 [入門編・理論編]" 朝倉書店
- [2] 橋間智博, "Web サイト訪問者の分類を行うアクセス解析ツール" FIT2004, pp. 413-416
- [3] 橋間智博, "IRT による情報提供型 Web サイトのアクセス分析" 情報科学技術レターズ, vol.5, pp. 423-425