

Web 教材データベースからの教材推薦サービスに関する研究

栗原 隆平*1

Email: sky-high-land@cup.ocn.ne.jp

*1: 東京電機大学大学院情報環境学研究科情報環境学専攻

◎Key Words 推薦システム, 協調フィルタリング, Web 教材

1. はじめに

今日、レコメンドシステムと呼ばれる機能の研究がさかんに行われている。レコメンドシステムとは、ユーザのコンテンツの閲覧履歴データなどをデータベースに保存し、ユーザが再び訪問または別のユーザが訪問した際、そのユーザに適切な、またはユーザの興味を引くことができるコンテンツをデータベースから割り出し、そのユーザにコンテンツを推薦・提供するものである。例えば、Amazon.com*ではユーザのデータから興味のある商品を割り出し、商品を推薦するシステムがある。Web 教材において学習コンテンツを推薦するシステムを適用すれば、ユーザに最適な学習コンテンツを提供することになる。これにより、ユーザの理解を深めたり、早めたりすることが可能であると考えられる。

教材の推薦方式として、図1に示しているように、推薦者がその Web 教材の中でどのようなコンテンツを閲覧してきたかなどの閲覧ログデータを用いて行われる。そのログデータを元にして被推薦者に傾向が類似している推薦者を選出し、推薦コンテンツを提供する流れになる。

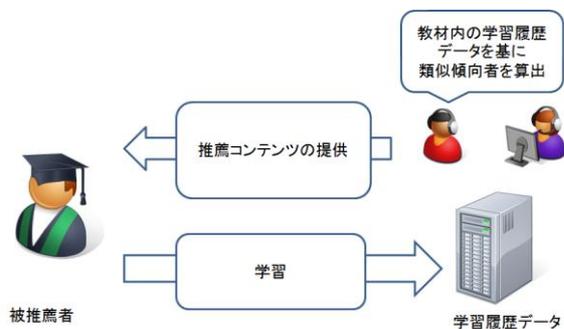


図1：教材の推薦

2. 既存研究と研究目的

2.1 既存研究

ログデータを用いたレコメンドは様々な場面で行われている [1][2][3]。しかし新規のユーザのように、サーバ側にログデータが無いユーザがアクセスしてきた場合、通常のリコメンド方式ではそのユーザの嗜好性などを把握・判断できないため、適切なコンテンツを提供することができないコールドスタート問題が懸念される。

2.2 研究の目的

本研究では、このコールドスタート問題に関連し、

ユーザの持っているログデータ以外の属性を用いてコンテンツを提供できるかを目的としている。今回はユーザの文章を読む速さと言う属性を軸に教材推薦を行えるか検証をする。

3. 検証

3.1 検証内容

本研究では、文章を読むスピードで理解しやすいコンテンツの種類（文章や簡単な図が多いコンテンツ・文章や簡単な図がメインだがもう少し詳しい図が加わったコンテンツ・FLASH形式で説明するコンテンツの3種類）に 差がでるかどうかを検証した。

3.2 検証方法

まず、被験者には自身の文章を読む速さの測定を行ってもらう。文字量が同じかつ内容の違う文章を3回読んでもらい、その平均を被験者の読むスピードと言う属性に設定する。

次に、本校の個別復習支援システム (AIRS) **で、20分間学習してもらった。学習してもらったコンテンツとコンテンツの種類は以下の通りとなっている。

- 1.文章や簡単な図が多いコンテンツ (Aパターン)
 - ・関数従属性
 - ・B+木へのkeyの導入
- 2.もう少し詳しい図が加わったコンテンツ (Bパターン)
 - ・直接編成ファイル
 - ・ビュー表の定義
- 3.FLASH形式で説明しているコンテンツ (Cパターン)
 - ・リレーショナルデータベーススキーマの導出
 - ・表の作成

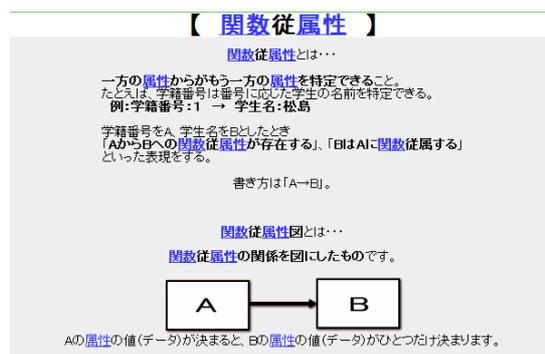


図2：コンテンツの例 (パターンA)

なお、このコンテンツの選出は筆者が無作為に抽出したものである。

20分間の学習後、5分間の休憩を取ってもらった後、学習したコンテンツの内容をそれぞれ説明してもらいどのコンテンツをどの程度理解しているかを測る理解度チェックを行った。この方法を用いた理由として、単純に用意された問題を解く場合に比べ学習内容を説明をする方が学習教材を理解していないと難しく、検証の信頼性が向上するからである。なお、理解度チェックの採点方法は10点満点の部分点方式である。

理解度チェック後、文章を読む速さと理解度チェックの結果を比較し、文章を読む速さでどの種類のコンテンツをよく理解しているか、文章を読む速さで理解しやすいコンテンツの種類に違いは出るのかを調べた。

3.3 検証結果

今回は本校のデータベースの講義を履修している学生 (Type A, Type B), 外国から留学してきている学生 (Type C), 情報系の分野の学習を行っていない学生 (Type D) の3種類の被験者を用意して検証を行った。

検証の結果、表1のように文字を読む速さが500文字以上/分の被験者は文章量が多いコンテンツの内容をしっかりと把握している傾向が出た。逆に文字を読む速さが遅い被験者は文章量が多いコンテンツを把握しきれず、うまく説明できなかつた結果となった。

表1：Aパターンでの比較結果

被験者	平均速度	A-1	A-2
Type A	511	10	0
Type B	626	7	2
Type C	309	2	0
Type D	883	8	2

一方、FLASH形式で表示されていたコンテンツに関しては、表2のようにすべての被験者が同等程度の説明内容である結果となった。

表2：Cパターンでの比較結果

被験者	平均速度	C-1	C-2
Type A	511	10	5
Type B	626	3	6
Type C	309	0	5
Type D	883	7	5

4. 考察

表1の結果からわかるように文章を読む速度が500文字以上/分の被験者はAパターンのコンテンツの内容をよく理解していることから、コンテンツを推薦する際、文章量が多いコンテンツを提供しても支障は少ないと言える。一方、読む速度が500文字に満たなかつた被験者は、文章量が多いコンテンツを提供した場合、内容を把握しきれない可能性がある結果となったので、文章量が多いコンテンツを推奨コンテンツとして提供

することは避けたほうが良いと言える。

表2からわかるように、CパターンのようなFLASH形式で表示されるコンテンツの場合、どの被験者もほぼ同等の理解度を示す結果となった。また、表3のように全体と比較した時も点数が高い傾向にあったことから、FLASH形式のような図が多いコンテンツは読む速度に関係なく高い効果が得られると言える。

表3：全体の配分

被験者	速度	A1	A2	B1	B2	C1	C2
Type A	511	10	0	2	6	10	5
Type B	626	7	2	0	6	3	6
Type C	309	2	0	5	3	0	5
Type D	883	8	2	4	0	7	5

パターンBのコンテンツに1問目に関しては、検証実験後に被験者から内容が把握しづらいコンテンツであったと言う意見が非常に多かった。また、2問目に関してはコンテンツの内容がデータベースに関する知識が前提にあるものとして考えられているコンテンツだったためこのような結果になったと思われる。これらのことはコンテンツを無作為に抽出したこともあるためと考えられる。

5. おわりに

今回の検証実験で、読むスピードと言う属性がWeb教材を推薦するうえで有効性がある可能性が得られた。

今回の検証実験では読む速度が500文字に満たない被験者が、日本語を母国語に持たない被験者のみであったことから、母国語が日本語でかつ読む速度が500文字に満たない被験者など様々なパターンの被験者を用意しこの検証を行い、さらなる確証を得ていく必要がある。

参考文献

- [1] 小野智弘 麻生英樹 本村陽一 “情報・コンテンツのレコメンド技術と課題” 電子情報通信会誌 Vol.94 No.4 (2011年4月) pp.310-315
- [2] 樽井 勇之 “協調フィルタリングとコンテンツ分析を利用した観光地推薦手法の検討” 上武大学経営情報学部紀要 36, 1-14, 2011-12
- [3] 松原 有希 長田 智和 玉城 史朗 “協調フィルタリングを用いた初等教育の授業向け Web 検索手法の提案” 情報処理学会研究報告. コンピュータと教育研究会報告 2007(101), 69-74, 2007-10-05
*:<http://www.amazon.com/> (2012年6月現在)
**:<http://airs.sie.dendai.ac.jp/> (2012年6月現在)

謝辞

本研究は、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 24500122 「ユビキタスデータベースにおける知識融合技術に関する研究」の支援による。