3U-5

顧客像の類似性をプロファイリング分析した書籍推薦システム

山村武司†，大野澄雄†，奥正廣†，千種康民†，小池隆‡
†東京工科大学，‡富士ソフト ABC

1 はじめに

現状で個人の書籍などの管理を最適に行う方法は確立されていない．そのため，個人の書籍などの管理を最適に行うためにデータベースを利用した個人蔵書管理のための Web サイトを構築し，パソコンや携帯端末から利用することにより，書籍の重複購入や買い忘れなどの問題を解決したいと考えた．

また，そのユーザの傾向や他ユーザの傾向や感想などの情報を分析し，そのユーザが興味を持つ新たな書籍を推薦できれば新刊の買い忘れをなくしたり，新たな書籍を読むきっかけになるのではないかと考えた．

amazon 等との違いとして，タイトルやレビューといった情報で，書籍を関連付けることにより，購買・アクセス履歴なしで，推薦することを可能とすることを目指した．

本研究では，昨年度「個人の蔵書情報を共有する書籍推薦サービスの研究」として行った成果を進め，特に，さまざまな手法がある書籍推薦において，顧客像の類似性をプロファイリング分析することによる効果を示す．

2 推薦システム

2.1 顧客の行動パターン

書籍を購入する場合，顧客の行動パターン，購入動機はいくつかに分けられる．

小説などの場合，これまで買っている著者，すでに購入しているシリーズもの，タイトルや表紙などの見た目といったものが重要な要素である．

専門的な書籍の場合，自分が興味を持っている分野の書籍，良いと思った出版社や出版シリーズなどが重要な要素である．

そして，よく売れている書籍や，友人知人による推薦をされた書籍はジャンルを問わず，購入動機になりやすい．

2.2 顧客像の類似性と推薦情報

顧客像の類似性を探し出すための手法として，特定ユーザ（本人）と他のユーザ（ユーザ A）の登録した書籍を比較し，同一書籍の所有率が高いほど類似度が高いと考える．(図 1)

その結果，本人が所有していない書籍を推薦する．

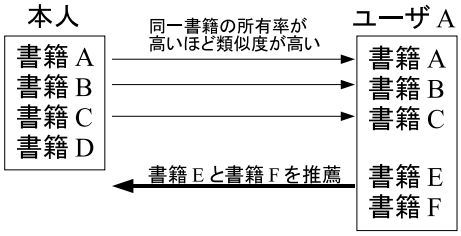

図 1: 類似性による推薦

2.3 蔵書情報と全ユーザの傾向を用いた推薦

まず，いくつもの単純なルールを重ね合わせることで，顧客の傾向を抽出する．これは，単純なルールを用いることで，従来，応答性が悪いとされるプロファイリング処理の高速化と高度化の両立を目指している．

推薦ルールは表 1 のようなものを採用した．

表 1: 書籍推薦ルール

ルール	ルール名	重み
1	出版社シリーズ（1 位〜3 位）	7・6・5
2	CCODE（1 位〜2 位）	7・6
3	CCODE1 桁目（1 位〜2 位）	2・1
4	CCODE2 桁目（1 位〜2 位）	2・1
5	CCODE3・4 桁目（1 位〜2 位）	2.5・1.5
6	出版社コード（1 位〜3 位）	6・5・4
7	新刊書籍	2
8	全ユーザ登録書籍ランキング	1
9	閲覧数	1
10	閲覧日時	1
11	登録日時	1.5
12	登録著者（1 位〜3 位）	7・6・5
13	タイトル単語	7
14	レビュー単語	7

2.3.1 本人の蔵書からの推薦

CCODE，CCODE1 桁目，CCODE2 桁目，CCODE3・4 桁目のルールを用い，まだ買っていない同一ジャンルの本の推薦を行っている．

出版社シリーズおよび出版社コードというルールで同一発行シリーズ，および，多く登録している出版社の本を推薦している．

2.3.2 全ユーザの傾向からの推薦

全ユーザ登録書籍ランキングというルールで，同一書籍を複数ユーザが所持している場合，推薦度合いの高い書籍であると考えたので推薦している．

閲覧数は，書籍の詳細データを閲覧した延べ回数が多い書籍は興味を引く書籍であると判断し，推薦を行う．

閲覧日時と登録日時は，書籍の詳細データを最後に閲

Books recommendation system based on customer profiles similarity
Takeshi YAMAMURA†,Sumio OHNO†，Masahiro OKU†，Yasutami CHIGUSA†，Takashi KOIKE‡
†Tokyo University of Technology, ‡FUJISOFT ABC, Inc.
E-Mail chigusa@cc.teu.ac.jp
URL:http://www.teu.ac.jp/chiit/~yamamura/BookProfiling/

覧された日時が新しいもの，および，書籍の詳細データが登録された日時が新しいものであり，それらは最近誰かの興味を引いた書籍であると判断し，推薦を行う．

新刊書籍とは，最近発売された書籍は興味を引くと考え，推薦を行う．

2.4 書籍の類似性による推薦

タイトル単語は，タイトル・サブタイトル・出版シリーズを単語で分割し，そのうち，名詞を抽出し，マッチングを行い，同傾向と考えられる書籍を抽出し，推薦を行った．(式 1)

$$\sum_{\text{単語 } 1}^{\text{単語 } n} (\text{書籍 } A \text{ の単語出現回数} + \text{書籍 } n \text{ の単語出現回数}) \qquad (1)$$

レビュー単語は，出版社から公表されているレビューを単語で分割し，名詞を抽出し，マッチングを行い，同傾向と考えられる書籍を抽出し，推薦を行った．(式 2)

$$\sum_{\text{単語 } 1}^{\text{単語 } n} (\text{書籍 } A \text{ の単語頻度} \times \text{書籍 } n \text{ の単語頻度}) \times \text{補正値} \qquad (2)$$

2.5 推薦手法の統合

複数のルール（表 1）をつくり，それぞれのルールに基づいてデータベースから 10 件の書籍を抽出する．そして，その書籍に対して抽出した 1 位から順に 10 ポイント〜1 ポイントという順位ポイントを付加している．また，そのときに，ルールごとの重みを設けており，順位ポイントに重みを掛けた合計がその書籍に対するそのルールでのポイントとなる．

そして，最後に全部のルールでのポイントを合計して，ポイントの高かったものが優先度が高いと考え，表示していく．(式 3)

$$\sum_{\text{ルール } 1}^{\text{ルール } 14} (\text{順位ポイント} \times \text{重み}) \qquad (3)$$

3 テスト用 Web サイト紹介

3.1 構造

本研究では，実際の書籍管理・販売を考慮したテスト用 Web サイトを構築し検証を行った．(図 2)

このテスト用 Web サイトは各表示部分をモジュールとして分割，および，機能を関数化することで，簡単に追加変更できるようにし，ユーザの簡便性や見やすさ，作成・保守のしやすさを最大限考えた．

また，そのことにより，複数のパターンのテストや新規のルールのテストを容易に行うことを可能にした．

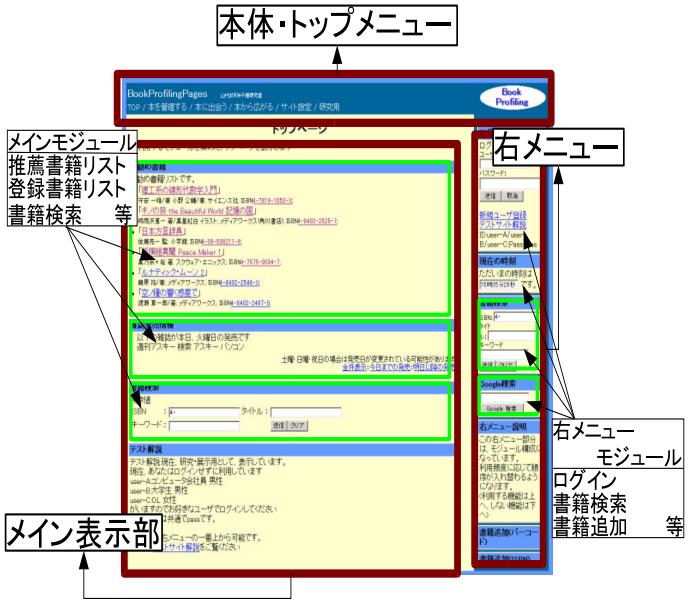

図 2: 基本全体画面

4 評価

単純なルールを複数採用することは，複雑なルールを構築することと比較し，簡単にルールの追加および修正を可能とした．また，プロファイルデータが十分蓄積されていないユーザや，まったく登録をしていないユーザであっても，一定の推薦を行うことを可能とすることが出来た．また，ルールを単純にすることにより，処理の時間を短くすることも可能となった．

なお，実際のユーザを登録し，最適な書籍が推薦されているかの評価を今後実施する．

5 まとめ

今後の課題として，書籍データが毎週膨大な量が追加されていくことにより，処理に時間がかかるようになることが予想されるため，より一層，データベースの構造や処理内容を工夫するなどの改良が必要になると思われる．

また，タイトル・レビュー等を用いて推薦するに当たり，単語を抽出しているが，推薦するときには役に立たない，あまりにも一般的過ぎる単語やすべての書籍を通して出現頻度が高すぎる単語などを含むと精度が極端に低下する事例が見られる．これを解決するために，推薦に使用しない単語を設けたりする必要がある．

参考文献

[1] 堀，小松，伊藤，新谷．エージェントによる書籍情報管理システム MiBook の試作，第 66 回情報処理学会全国大会講演論文 (2)，pp347-348，2004.