

書籍情報管理システム MiBook における 書籍の利用状況に基づく推薦機構について

About recommendation mechanism based on the utilization status
in book information management system MiBook

堀 和裕*¹
Kazuhiro HORI

小松 正樹*¹
Masaki KOMATSU

伊藤 孝行*¹
Takayuki ITO

新谷 虎松*¹
Toramatsu SHINTANI

*¹名古屋工業大学大学院 工学研究科 情報工学専攻
Graduate School of Engineering Nagoya Institute of Technology

We developed a book information management system MiBook. MiBook has several support functions that support users to acquire knowledge from books, such a book recommendation function. MiBook can recommend a book from similarity between interests of the user and the book data. However, contents of the recommended book often has been already learned by the user. Therefore, in this paper, we propose a user's recommendation mechanism based on the utilization status of books, and propose a model of knowledge acquired from the book.

1. はじめに

本論文では、書籍情報管理システム MiBook における書籍の利用状況に基づく推薦機構について述べる。書籍は知識を獲得する一般的な手段である。研究室等の、求める知識が近いコミュニティではどのような書籍を誰が所有しているかという情報の管理は有益である。登録された書籍情報からユーザプロフィールを作成し、ユーザに書籍を推薦することで書籍から知識を獲得する機会を増やすことができる。そこで我々の研究室では、書籍による知識獲得の機会を向上させるために書籍情報管理システム MiBook の開発を行っている。

MiBook ではユーザ毎の書籍情報の管理、書籍貸借管理、および書籍推薦を行う。書籍情報は Amazon Web Services(以下, AWS とする) [Amazon Web Services] を用いて Amazon データベースより情報を取得する。AWS で必要な情報が得られない場合は Amazon の Web ページにアクセスして必要な情報を取得している。

我々は、以前の研究において書籍推薦をユーザプロフィールおよび書籍データの類似度を用いて行った [堀 04]。ここでは、協調フィルタリングなどに一般的に用いられるユーザプロフィールと類似度の高い書籍をユーザの興味ある書籍として推薦する手法を使った。

以前の研究や Amazon での推薦手法は、ともにユーザの興味を基に推薦している。これらの手法では、ユーザに興味のある書籍の推薦には優れている。しかし、書籍をユーザが新たな知識を獲得するための媒体とした場合には、必ずしも十分な推薦が提供されるとはいえない。例えば、推薦された書籍を実際に利用した場合に、内容の多くは既に十分な知識を習得したものであり、ユーザの興味ある分野に関する書籍の推薦という点では十分だが、ユーザにとって知識の習得に役立つ書籍の推薦という点では十分でない。また、ユーザにとって新しい分野の知識を習得する場合に、書籍の推薦が発生しないという問題もある。例えばプログラミング言語 perl について新たに知

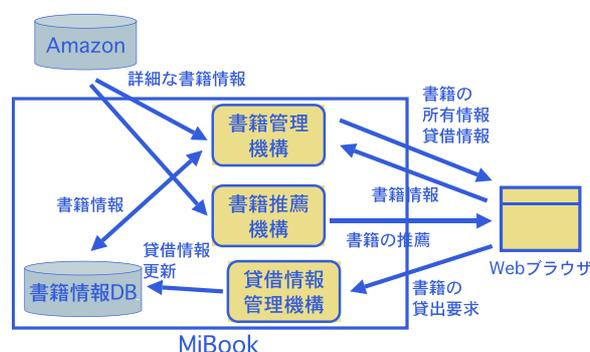


図 1: MiBook システム概要

識を得る場合、多くの場合、検索により書籍の選定を進めていく。しかし perl の初心者向け書籍がどれなのか分からないため、どの書籍を読めばいいかを確認していかなければならず負担が大きい。そこで、本論文ではユーザが書籍より得た知識に注目し、ユーザがある分野において何の知識を持っているか、何の知識を持っていないかという基準より推薦を行う推薦手法を提案する。

本論文の目的は、MiBook での書籍推薦において、ユーザの習得している知識を基にユーザがより多くの知識を得られる書籍を推薦する推薦機構の提案である。ユーザの知識は MiBook を通してユーザが利用した書籍の情報よりモデル化していく。

本論文では、2. 章で書籍情報管理 MiBook と主な機能について述べる。3. 章で今回提案するユーザの利用状況を形成するモデルについて述べ、4. 章で利用状況に基づいた書籍推薦の手法について述べる。5. 章で考察をおこない、6. 章で関連研究について述べる。7. 章で本論文をまとめる。

2. 書籍情報管理システム MiBook

MiBook は本研究室で開発された論理型記述言語 MiLog[Fukuta 01] のサーバー機能を用いて実装された Web アプリケーションである。本システムは Web ブラウザを介して利用する。図 1 に MiBook のシステム概要を示す。

連絡先: 堀 和裕, 名古屋工業大学知能情報システム学科新谷研究室, 〒 466-8555 名古屋市昭和区御器所町, TEL: 052-733-6550, FAX: 052-735-5584, E-mail: kazuhiro@ics.nitech.ac.jp

MiBook にはユーザ書籍情報管理機構、書籍貸借管理機構、および書籍推薦機構が存在する。以下、本章では三機構について述べる。

ユーザ書籍情報管理機構

ユーザ書籍情報管理機構ではユーザ情報と書籍情報をあわせてユーザ書籍情報の入力、表示、および管理を行う。ここで、書籍情報とは書籍のタイトル、著者、出版社、ISBN、出版日、目次、Amazon 上でのジャンル、書籍提供側のレビュー、および書籍購入側のレビューである。ユーザ情報とはユーザの書籍に対する評価、レビュー、書籍の所在、および利用形態である。利用形態とは対象となる書籍を所有しているのか、他ユーザより借りた書籍なのか、という情報である。

図 3 はシステムに登録された書籍情報の一覧表示である。タイトルのリンクより詳細情報を閲覧できる。

本システムでは、ユーザの所有している書籍を登録する際に、ISBN、書籍名、著者名、出版社名、出版日、および、書籍に対するユーザの評価、書籍に対するユーザのレビュー、および書籍の所在を入力する。ISBN が入力されている場合は必要その他の情報が直ちに取得される。全ての入力をユーザに要求するのは負担が大きいので、一部の入力情報を基に書籍の一般的な情報を取得可能とした。書籍情報として ISBN が入力されていない場合は、候補となる書籍を表示し、ユーザが自分の登録すべき書籍を候補の中より選定する。取得対象となるのは ISBN、書籍名、著者名、出版社名および出版日である。ユーザの評価およびレビューは他ユーザが対象書籍を選定する場合の判断基準になる。ユーザの評価はユーザプロフィール作成時に対象書籍を用いるかを判断する基準となる。

書籍貸借情報管理機構

書籍の所有者が自分ではなく、かつ対象となる書籍を所有者が他のユーザに貸さずに管理している場合、貸借管理エージェントを通して書籍の所有者に書籍の貸出要求の送信が可能である。図 3 の貸出の列にある「借りる」のリンクより実行される。貸借管理エージェントは貸出要求を受け取った場合、所有者に対して書籍貸出の可否を訪ねて結果をユーザに返す。所有者と貸借エージェントはメールを介して情報の受け渡しを行う。対象書籍を所有者が他ユーザに貸している場合や、所有者自身が使用中である場合は貸出要求を出す代わりに貸出希望を書籍の所有者に対して送る。この際、書籍の所有者は誰がどの本の貸出を希望しているかをメールで受け取る。

書籍を借りた場合は一定の期間が経過するとエージェントより所有者への書籍の返却要求が行われる。エージェントは書籍の返却時に書籍を借りたユーザに借りた書籍に対するユーザの評価を要求する。貸借情報は評価と共に履歴として保存され、ユーザプロフィールの作成時に利用する。

書籍推薦機構

本研究では書籍情報として Amazon の記述したレビュー、出版社の内容紹介、BOOK データベース^{*1}もしくは MARC データベース^{*2}に格納されている内容紹介および購入者の記述したレビューより存在する情報を組み合わせて用いる。Amazon の記述したレビューと、出版社の内容紹介および購入者の記述した

*1 トーハン、日本出版販売、紀伊国屋書店、日本アソシエーツの 4 社で共同構築している日本国内出版図書の内容情報などを収録したデータベース

*2 世界の 20 数カ国の国立図書館において、それぞれ収集した出版物の書誌情報などを電子化したデータベース

タイトル	著者	出版日	所有者	貸出
Artificial Intelligence (3rd Edition)	Winston	1992/01	toralab	借りる
Artificial Intelligence: A Modern Approach (2nd Edition)	Stuart J. Russell, Peter Norvig	2002/12	toralab	借りる
Essential JNI: Java Native Interface	Rob Gordon	1998/03	kazuhira	-
Javaによる知能プログラミング	新谷 虎松	2002/02	kazuhira	-

図 2: MiBook に登録された書籍の表示

たレビューは情報の有無が書籍によって違うので、複数から取得する方法を採用している。購入者の記述したレビューは対応する書籍についての内容が記述されているかの信憑性を計るために、他の人が購入者のレビューに対して賛同したかをとという情報をレビュー用いる判断基準とする。以上の情報は AWS を用いて主に取得するが、BOOK データベース、MARC データベースおよび購入者のレビューに対する他者の賛同は AWS で取得ができない。そこで、AWS では取得できない上記の情報については、AWS にある書籍情報の URL を基に Web ページにアクセスし、Wrapper を用いて必要な情報を抽出する。

以前の研究では、書籍の推薦はプロフィールを今まで利用した書籍の情報より語の共起関係を基に作成し、推薦候補の書籍情報との類似度計算より推薦書籍を決定した。しかし、推薦された書籍を検討した所、ユーザの興味を反映しているが、既に習得している知識しか内容に含まれていないため、あまり役に立たなかったという問題が報告されている。

3. ユーザの利用状況のモデル化

本章では、書籍情報より対象書籍に含まれる知識を表現する手法について述べる。ユーザが MiBook 内で利用した書籍の書籍情報より以下の手順で表現する。始めに MiBook 内に格納されている書籍データより書籍のタイトル、および書籍の目次を取得する。次に書籍のタイトル、および書籍の目次より名詞を抜き出し、リストで格納する。書籍の目次は各章毎にリストで格納する。同一リスト内で単語が重複した場合は一つだけリストに格納する。即ち、一つの書籍に対して一つの書籍タイトルのリストおよび複数の目次のリストが存在する。最後に、書籍タイトルの名詞リストを親ノード、書籍の目次より抽出した名詞を子ノードとした木構造を形成する。親ノードは対象書籍に含まれる知識の概要を表現し、子ノードは対象書籍に含まれる具体的な知識を表現しているとする。例として新谷虎松 著「Java による知能プログラミング入門」(コロナ社)[新谷 02]の書籍タイトルおよび目次を図 3 に、名詞を抜き出してリストとし、木構造化したものを図 4 に示す。木構造を用いているのは、書籍タイトルと目次が書籍の内容を表す上で階層構造になると考えたためである。

次に木構造化したデータをカテゴリ毎に分けて格納する。今

タイトル：Javaによる知能プログラミング入門

目次

- 1 Javaの基礎(プログラムの作成と実行 Javaの基本的な文法 アプリケーションとアプレット マルチスレッド 簡単な電卓プログラム)
- 2 探索とパターン照合(探索パターン照合)
- 3 知識表現と推論システム(知識表現 ルールベースシステム プランニング 整合性管理機構)
- 4 エージェントと意思決定支援技術(モバイルエージェント 意思決定支援システム)

図 3: 書籍のタイトルと目次

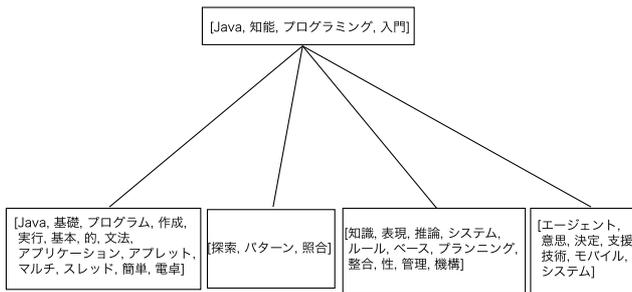


図 4: 書籍のタイトルと目次の木構造化

回はカテゴリとして Amazon がジャンル別として分けているデータより最下層のデータをカテゴリとした。例えば上述の「Java による知能プログラミング入門」では以下のジャンルに分けられている。

ジャンル別 > コンピュータ・インターネット > インターネット・Web 開発 > Web 開発 > JAVA

この場合、「Java による知能プログラミング入門」のカテゴリは JAVA になる。

今回はユーザの評価が高い書籍に対して、対象ユーザが書籍内の知識を習得していると見なす。よって、ユーザの利用書籍より、ユーザの評価が高い書籍全てに対して以上の処理を行う。カテゴリではカテゴリ名を親ノード、書籍のタイトルの名詞リストを子ノードとした木構造を形成する。名詞をリストとして扱うのは、一単語で区切ると「入門」といった他の単語との組み合わせによって意味の出る語を考慮したためである。また、書籍の中には複数のジャンルに登録されている書籍も存在する。この場合は、各ジャンルを基に作られるカテゴリ全てに登録する。即ち、カテゴリ間で重複する書籍情報が存在する可能性がある。以上の処理より各カテゴリ毎に導出される木構造を本研究における書籍の利用状況モデルとする。

4. 書籍の利用状況に基づく推薦

本章では、ユーザが今までに獲得していない知識を多く含む書籍を推薦する手法について述べる。前章で述べた書籍の利用状況モデルを用いて以下の手順で推薦を行う。1、カテゴリ毎に推薦候補となる書籍のデータを取得する。本稿でのカテゴリは、Amazon 上のジャンルに相当するため、AmazonDB よりジャンル毎の書籍のデータを取得する事になる。2、取得し

た推薦候補の書籍データよりユーザが既に利用した書籍を取り除く。3、推薦候補の書籍データに対して 3. 章の木構造化処理を行う。4、推薦候補の各書籍とユーザの対象カテゴリにおける利用状況との類似を比較し、対象書籍より新たに獲得できる知識の程度を算出する。5、算出された値を元に推薦する書籍を決定する。以上の処理を各カテゴリ毎に行い、カテゴリ毎に推薦書籍を導出する。

類似の比較とは二つのデータ間で同じ語がどれだけ含まれているかであるとする。推薦候補の書籍データとユーザの対象カテゴリにおける利用状況の類似の比較は、書籍の目次にある各章の類似を比較し、どの目次が類似しているかを算出する。一方の章が他方の章と類似している場合、両章より獲得される知識は同一であるとする。目次の章で類似の比較を行う前に、書籍のタイトルで一度二つの書籍の類似の比較を行う。二つの書籍のタイトル間の類似が低ければ、二つの書籍の内容は異なるものである可能性が高いと考えるためである。推薦候補の書籍に含まれる知識とユーザの書籍利用状況モデルに内在する知識を比較し、書籍利用状況モデルにない知識が推薦候補の書籍に多く含まれていれば、推薦候補の書籍はユーザにとって新たな知識を多く含んだ書籍であるとする。

類似の比較は式 (3) で行う。 T_x, T_y は図 4 に見られるリスト構造を持った情報源である。

$$com(T_x, T_y) = \frac{\sum_{i=1}^{L(T_x)} \sum_{j=1}^{L(T_y)} f(x_i, y_j)}{\min(L(T_x), L(T_y))} \quad (1)$$

$$f(x, y) = \begin{cases} 1 & (x = y) \\ 0 & (x \neq y) \end{cases} \quad (2)$$

x_i, y_j は情報源 T_x, T_y におけるリストの i 番目、 j 番目の要素、 $L(T_x)$ はそれぞれは情報源 T_x のリストの長さを表す。 $f(x, y)$ は x と y が一致する場合は 1 を、一致しない場合は 0 を返す (式 (4))。

目次間の類似の比較は、 C_X を推薦候補の書籍の目次の集合 C_Y を利用した書籍の目次の集合とすると、推薦書籍の目次で既に獲得した知識が含まれている割合は式 (5) で表される。

$$content(C_X, C_Y) = \frac{\sum_{i=1}^{L(C_X)} r(X_i, C_Y)}{L(C_X)} \quad (3)$$

$$r(X, C_Y) = \begin{cases} 1 & (\sum_{i=1}^{L(C_Y)} com2(X, Y_i) \geq 1) \\ 0 & (\sum_{i=1}^{L(C_Y)} com2(X, Y_i) < 1) \end{cases} \quad (4)$$

$$com2(X, Y) = \begin{cases} 1 & (com(X, Y) \geq \alpha) \\ 0 & (com(X, Y) < \alpha) \end{cases} \quad (5)$$

X_i, Y_i は目次 C_X, C_Y の i 番目の要素である。式 (7) は類似の比較が一定値を越えた場合に同一の知識を含む章であるかを調べる。式 (6) では目次 X が目次の集合 C_Y と一つ以上同一の知識を含むかどうかを調べる。以上の式より推薦候補の書籍に対してどれだけ既に獲得した知識があるかを算出する。式 (5) より本論文では推薦候補の書籍を三つに分ける。新たな知識を多く含む書籍、新たな知識と既に得ている知識が同程度ある書籍および既に得ている知識を多く含む書籍である。今回は新たな知識を多く含む書籍より推薦書籍を決定する。

5. 考察

従来の推薦の多くはユーザの興味に注目し、ユーザの興味に近いものを提示している。即ち今まで読んだ書籍を基にして興味に近い書籍を新たに推薦している。本論文では、ユーザが書籍より得られた知識に注目している。ユーザが何を知識として持っており、何を知識として持っていないかを明らかにする事で、ユーザが次に習得すべき知識について述べている書籍を推薦している。研究活動を行う上で、書籍は基礎的な技術の知識や専門分野の知識を得る対象と捉える事ができる。従来の推薦手法を書籍に用いた場合、ユーザが今までに読んだ書籍より関連度の高い書籍を推薦している。この方式では、今までに読んだ書籍と推薦される書籍の間に関連する知識と関連しない知識を考慮してない。基礎的な技術の知識は、より多くの知識を得られる書籍を読む方が効用が高いと考えられる。よって、書籍情報よりユーザの習得した知識を基にする本論文の推薦手法は有用であると考えられる。

さらに、MiBook ではコミュニティ内のメンバーの書籍情報を管理している。あるユーザが獲得していない知識について、他メンバーの獲得している知識と書籍の利用状況を照合することで具体的にどのような知識を獲得していないかの判別が可能である。予めユーザが知らない知識を明示できる事で、推薦書籍を選定する段階においてユーザの獲得していない知識を含んだ書籍を選定する、といった処理が可能となる。ユーザが獲得している知識を多く含む書籍を推薦候補より取り除く処理と並列に進める事で効率よく推薦候補の絞り込みが可能になると考えられる。特にユーザが新たな分野についての書籍を欲する場合、他メンバーの書籍情報より対象となる分野での共通する知識より推薦書籍を選定できる。ある分野において他メンバーが共通で持つ知識は対象分野において必要となる知識である、と考える事ができる。即ち、新たな分野の知識を得る場合にも、有用な書籍推薦が可能であると考えられる。

今回提案した推薦手法と以前の研究で用いた提案手法は相補的な手法である。新たな書籍を読む場合、自分の興味を基に選ぶか、新たな知識を得る書籍を選ぶかで基準が大きく異なる。選択的に推薦手法を使い分ける事で状況に応じた推薦が可能となる。

6. 関連研究

書籍を扱う研究としては幾つかの研究が進められており、書籍情報の取得、仮想図書館の構築、書籍情報を用いた推薦、および検索において報告がある。[寸田 02] では、Web 上の形式の異なる情報を自動収集・統合し、選書作業を支援するシステムを構築している。[Mooney 00] では、情報抽出および機械学習による敏捷のカテゴリ化を用いた書籍の内容を基にした推薦を行う。[箱崎 98] では、仮想図書館を構築している。個人利用への最適化を目的とし、仮想司書エージェントにユーザの個人情報管理させている。また、[白木 00] では図書検索の精度を高めるために検索式と目次タイトルとの柔軟な照合を行っている。その際、目次の階層的構造に注目している。いずれの研究においても書籍の利用状況を基に書籍より獲得した知識を考慮しているものはない。図書目録のオンライン検索システム OPAC(Online Public Access Catalogue) に見られるように、既存の図書館情報システムでは図書検索を目的としたシステムがほとんどである。我々の研究や [箱崎 98] のように、利用書籍の管理といった個人情報を扱ってはいない。不特定多数

の利用者に書籍情報のサービスを提供するのではなく、コミュニティ内のユーザにとって利用しやすい環境の構築を目指している点が異なる。

7. まとめ

本論文では書籍情報管理システム MiBook における書籍推薦としてユーザの書籍利用状況に基づいた推薦機構について述べた。ユーザの書籍利用状況をモデル化する手法を提案し、ユーザの書籍利用状況に基づいた推薦を行っている。

ユーザの書籍利用状況をモデル化する事で、ユーザが書籍より得た知識を明らかにする事が可能となる。ユーザが今までに習得している知識、習得していない知識の明示により、習得していない知識について述べている書籍を新たに推薦書籍として提示している。

今後の展開としては、モデル化された書籍利用状況を用いてユーザの知識の視覚化、およびどのユーザが何を知っているか、という検索機能の付加が考えられる。また、本稿の内容とは関連が薄いですが、MiBook と他の Web ページをシームレスに利用するために、MiLog にあるプロキシ機能を利用したブラウジング中での情報提供を考えている。オンライン書店で MiBook 内に登録された書籍に関するページを閲覧した際、付加情報として MiBook 内の情報を表示させる。

参考文献

- [堀 04] 堀和裕, 小松正樹, 伊藤孝行, 新谷虎松” 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集第 2 分冊 pp.347-348, 2004
- [Fukuta 01] N. Fukuta, T. Ito, and T. Shintani, ”An Approach to Building Mobile intelligent Agents Based on Anytime Migration”, Lecture Notes in Artificial Intelligence, Vol.2112, pp.219-228, 2001.
- [Mooney 00] Mooney R. J. and Roy L., ”Content-Based Book Recommending Using Learning for Text Categorization”, Proceedings of the V ACM Conference on Digital Libraries, San Antonio, USA, p195-204, 2000.
- [白木 00] 白木伸征, 黒橋禎夫” 自然言語入力と目次との柔軟な照合による図書検索システム” 情報処理学会論文誌, Vol.41, No.4, p1162-1170, 2000
- [寸田 02] 寸田五郎, 江口浩二” Web 上の情報資源を利用した選書支援システム” 情報処理学会第 64 回全国大会講演論文集, Vol.3, pp.125-126, 2002
- [箱崎 98] 箱崎勝也, 金井秀明, 石川克則, 陳泓, 井澤克司” 個人利用に適合した仮想図書館の構築” 電子情報通信学会論文誌 D - II, Vol.J81 D II, No.5 pp.1025 - 1034, 1998
- [新谷 02] 新谷虎松” Java による知能プログラミング入門” コロナ社, 2002
- [Amazon Web Services] Amazon Web Services
<http://www.amazon.co.jp/exec/obidos/subst/associates/join/webservices.html>