

貸出記録を用いた図書推薦システムにおける重みづけの変更

A trial approach of weighting for library loan records for
developing a book recommendation system

原田隆史, 増田浩佑

慶應義塾大学文学部

〒108-8345 東京都港区三田 2-15-45

E-mail : ushi@slis.keio.ac.jp

図書館で貸出記録を用いて適切な図書の推薦を行うために、貸出データをどのように取り扱うことが有効なのかを検討した。具体的には、大学図書館の貸出記録 272,751 件および推薦図書を評価する被験者 10 名の貸出記録に対し、「図書の分野」や「時間（貸出日付）」といった基準から全 5 パターンの重みづけを行い、これと協調フィルタリングにより図書の推薦実験を行った。実験の結果、貸出記録をそのまま用いた推薦では有用と評価された図書が平均 45.5%であったのに対し、重みづけを行ったパターンによる推薦図書はいずれも平均で 65%以上が高い評価を得られた。特に、重みづけに加えて貸出記録をさらに複数に分割して用いたときに、利用者にとって関心のある図書をより多く推薦できることが明らかとなった。

For the purpose of developing the book recommendation system, we have explored how to utilize library loan records effectively. 5 patterns of weighting were given based on the attributes such as "classification of books" or "checkout dates" to 272,751 loan records of the academic library and the loan records of 10 test subjects who evaluated the recommended books. The experiment was then conducted using these records and collaborative filtering. The result showed that more than 65% of the recommended books of every pattern of weighting obtained favorable evaluations, compared to 45.5% of those without weighting. Especially, the system was more effective to recommend appropriate books for users in case the loan records were categorized into several groups in addition to weighting the records.

1. 図書の推薦システムへの期待と貸出記録

現在、インターネットは国内における利用者が 9000 万人を超えたといわれており¹⁾、いまや日常生活に密着した主要な情報源として定着している。しかし、インターネットの普及や Web の発達、一方であまりにも情報が増えすぎるといって「情報過多」を深刻化させ、人が自分の望んでいる情報を自力で見つけることをますます困難にしている。

このような問題への対策として、近年、膨大な情報の中から利用者のニーズに適したものだけを選出し、利用者へ推薦するレコメンドサービスが注目されている。レコメンドサービスは、

行動履歴や情報への評価などにもとづいて各利用者の嗜好を推測し、個人に適合した推薦を行うことができるサービスである。様々な性質の情報に対応するため、これまでに数多くの手法による推薦システムが提案され、オンライン書店の Amazon をはじめ多くの電子商取引や情報提供サービスにおいて実用化されている。そして、効率的な情報の獲得やサービスの活性化といった利点から、情報の利用者、提供者双方にとって非常に有益な役割を果たしている。

そうしたレコメンドサービスの利点から、図書館においても、推薦システムを用いて積極的に利用者へ図書を推薦できないかという期待が

高まっている^{2) 3)}。その契機となったのは、OPACの機能や操作性への見直しである。図書館にとってOPACは、利用者と蔵書とをつなぐ重要なツールであるが、複雑な検索ロジックを理解しなければ使いこなせない点や、自由なキーワードによる検索が難しい点などが欠点として指摘され⁴⁾、GoogleやAmazonなどの優れたWebサービスに比べて後れをとっていることが問題点として指摘されてきた。近年、このような問題点に対する解決策として、利用者志向の次世代OPACの検討も急速に進められてきている。レコメンドサービスも、その機能の一つとして注目されている。

図書館でレコメンドサービスを展開できるかということについては、岡本真がその可能性を論じている⁵⁾。岡本は、Amazonでの推薦に用いられる購入記録に相当するものとして図書の貸出記録を挙げており、これを活用することができれば、図書館はAmazonと同等以上のレコメンドサービスを実現できると主張している。その根拠として、2005年の取次ルートによる書籍の販売点数が約7億4000万冊⁶⁾であるのに対し、同年の公共図書館の貸出点数は合計で約6億1000万冊⁷⁾であることを引き合いに出し、両者が規模の上でも大差がなく、利用者の行動履歴というデータとして利用可能ではないかという指摘を行っている。Amazonが書籍の販売のためにレコメンドサービスを非常に効果的に活用していることを踏まえれば、図書館においても、貸出記録を用いたレコメンドサービスを検討することは自然な流れといえよう。



第1図 Amazonの「おすすめ商品」サービス

しかし現状では、貸出記録をレコメンドサービスのような積極的な目的で収集・活用することには消極的な図書館も少なくない。これは、貸出記録には利用者が何を読み、興味を持ってきたのかを示す個人情報としての側面もあり、その保存自体が利用者のプライバシーを侵害することにつながりかねないという懸念に基づくものである。そのため、公共図書館を中心とした多くの図書館では、資料の返却後には貸出記録を消去している⁸⁾。

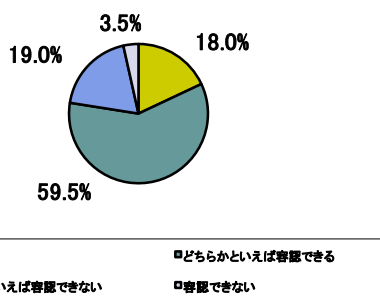
このような貸出記録を保存しないという態度に対しては、疑義を表明する意見も近年になって多く聞かれるようになってきている。たとえば、渡邊斉志は、Amazonの「おすすめ商品」機能を例として、利用者の行動履歴を蓄積・活用したサービスが多く顧客に受け入れられているのにもかかわらず、図書館は「利用情報を保持しない」という行動準則により、新たなサービスの開発・普及を停滞させてしまったと考察している⁸⁾。また、安東奈穂子らは、図書館利用者の信頼がプライバシー保護の遵守のみでは十分に得られるものではないとし、利用情報を活用したサービスが利用者の真の利益につながるとまで主張している⁹⁾。

さらに、現在でも貸出記録を実際に活用している例がある。いくつかの大学図書館では貸出記録を収集しており、これを分析することによって図書館における一定の利用動向を把握し、図書館の経営やサービスの方針を決定するのに役立てている。たとえば、星野雅秀ら¹⁰⁾は、東京大学総合図書館の一年間の貸出記録を分析し、学生の専攻と利用図書との間に関連性があることを明らかにしている。また、衛藤広隆ら¹¹⁾は、岡山理科大学図書館の貸出統計データの分析から年間や学年別の図書の利用傾向を明らかにし、選書方針などを打ち立てている。

以上の例にみられるように、一見多様な目的をもって利用されると考えられがちな図書館にあっても、貸出記録を収集して活用すれば、利用行動に一定の傾向を見出すことができると考えられる。これは、貸出記録が図書館利用者の行動や興味を把握するために確かに有用であり、

これが資料の管理という目的以上の価値を持つということを示しているといえよう。このような点からも、貸出記録を消去することは、利用者の潜在的な情報ニーズを示唆する貴重なデータを放棄するということにつながりかねないといえよう。

また、貸出記録を用いてレコメンドサービスを行うことの是非については、佐浦敬之が図書館利用者400名に対して意識調査を行っている¹²⁾。佐浦の調査によると、まず、新しいサービスのために貸出記録を保存することについては、利用者全体の約8割が容認できると回答し、貸出記録を用いたレコメンドサービスを使ってみたいかという質問に対しては、過半数の利用者が使ってみたいと回答したことが明らかとなっている(第2図、第1表)



第2図 新しいサービスのために貸出記録を保存することの是非 (n=400)

第1表 他人の貸出記録を使ったレコメンドサービスに対する意向調査(図書館利用頻度別)

	全体	とても使いたい	どちらかといえば使いたい	どちらかといえば使いたくない	全く使いたくない
全体	400	47	164	128	61
	100.0	11.8	41.0	32.0	15.3
週に1回以上	100	17	40	28	15
	100.0	17.0	40.0	28.0	15.0
月に1~3回	80	14	35	25	6
	100.0	17.5	43.8	31.3	7.5
年に数回	70	10	29	24	7
	100.0	14.3	41.4	34.3	10.0
全く利用しない	150	6	60	51	33
	100.0	4.0	40.0	34.0	22.0

第2図および第1表に見られるように、図書館利用者の多くは、図書館側が懸念しているよりも貸出記録を保存・活用することに寛容であり、従来の図書館の貸出記録を返却と同時に削除すべきという意識と、利用者の意識との間には隔たりがあることがわかる。利用者らは図書館が貸出記録を収集・保存する際に、あくもで

もプライバシーの遵守やサービス利用の自由などといった条件を求めているものの、個人情報を提供するリスクを許容し、新しいサービスを楽しむことを期待しているのである。佐浦の調査は、図書館におけるレコメンドサービスへの潜在的な利用者が存在することを明らかにしていると同時に、図書館側が貸出記録への認識を改め、新しいサービスへの活用も視野に入れた、貸出記録の取り扱い環境を整えていくことが必要であるということを示しているといえよう。

本論では、従来行われてこなかった貸出記録を用いたサービスの展開を視野に、情報推薦の手法のひとつである協調フィルタリングの手法を用いて図書館という条件下で Amazon のようなレコメンドサービスを行うことの有効性について考察を行うものである。特に、図書館に適応した推薦システムを実現するには、システムとしてどのような性質のものを用いればよいか、また、貸出記録をどのように扱えばよいか、検証することを目的としている。

2. 情報推薦システムの手法と図書館への適応

2.1 情報フィルタリング

多くの情報推薦システムで用いられているのが、情報フィルタリングと呼ばれる技術である。情報フィルタリングとは、利用者の興味や関心を記したプロフィールを参照して、情報源から次々と流れてくる情報のうち利用者の関心があるものだけを取り出すことをいう¹³⁾。プロフィールは、利用者の直接入力した嗜好情報、過去の行動履歴、あるいは対象とする情報や商品(以下、アイテム)の内容情報などのあらゆる情報をもとに作成され、これをもとに、利用者同士や、利用者の興味のあるアイテムと他のアイテムとの関連性の高さを予測し、個人に特化した推薦を行うことが可能である。

情報フィルタリングで用いられるアルゴリズムは、プロフィールの作成や推薦スコアを推測する時に利用する情報によって、内容ベースフィルタリング、協調フィルタリング、およびハイブリッド型フィルタリングに大別される。以

下では各方式の概要を述べ、図書館での事例を中心に、それらを用いて書籍の推薦を行ういくつかの試みについて考察する。

2.2 内容ベースフィルタリング

内容ベースフィルタリングは、利用者のプロフィールにみられる興味表現と、アイテムの持つ内容情報とを比較することによって推薦情報を決定する手法である¹⁴⁾。すなわち、システムは利用者が高い評価を与えたアイテムの内容と類似した内容を持つ他のアイテムを選出し、利用者に提示する。この方式では、推薦対象の内容自体を参照するため、アイテムや利用者の特徴表現が得やすく、蓄積された利用者やアイテムに関するデータが少ない場合でも、比較的適切な推薦を行うことができる。一方、特徴表現の記述はキーワードなどのテキスト情報に限定されるため、内容的に類似したアイテムばかりが推薦されやすく、意外性のあるアイテムを提示することは難しい。

廣岡康雄らは、購買履歴と書籍情報を用いて顧客ごとに適切な推薦情報を自動生成する内容ベース方式の書籍推薦システムを提案し、オンライン書店へ適用する実験を行っている¹⁵⁾。これは「顧客情報を表す顧客プロフィール」と「書籍の特徴」をそれぞれキーワード・ベクトルによって表現し、それらの間の類似度を計算することで書籍を推薦する。また、顧客のキーワード・ベクトルは書籍のカテゴリごとに別々に生成し、それらと書籍情報との間できめ細かに比較を行うことで、意外性のある書籍を推薦できるような工夫も凝らされている。このシステムは図書館での導入を前提としたものではないが、語の出現頻度にもとづいた手法という点では書籍一般に適用し得るものであり、図書館においても書誌情報を用いれば同様な推薦が行える可能性がある。

2.3 協調フィルタリング

協調フィルタリングは、アイテムへの評価や購買記録から得られる利用者の嗜好情報をもとに類似した嗜好を持つ利用者をグループ化し、対象利用者がまだ知らず、グループの他の利用者が高く評価しているアイテムを推薦していく

手法である¹⁴⁾。アイテムに関して必要な情報はその識別番号のみであり、利用者同士の関係によって推薦対象を決める点にその大きな特徴があるといえる。そのため、推薦アイテムがその内容による制約を受けず、思いがけないアイテムを推薦することも可能である。ただし、これまでに利用されず、評価もされていないようなアイテムは推薦対象とはならず、特に新着アイテムは対象外となりやすい。また、この方式を有効に機能させるためには、アイテムに対して十分な利用者の嗜好データが必要である。協調フィルタリングは Amazon においても採用されており¹⁶⁾、数千万規模の利用者と商品数を持つ Amazon は、この手法を最大限活用できる環境を有していると考えられる。

図書館の推薦システムとして協調フィルタリングの使用を検討している例としては、原田隆史の研究が挙げられる¹⁷⁾。この研究では、図書館の推薦を受ける被験者の貸出記録、および大学図書館に蓄積された貸出記録をもとに協調フィルタリングの手法で推薦図書を選出し、それを各被験者に提示した結果、一定の評価を得られている。ただし実験では、推薦図書の中に被験者の貸出と関連のあるものだけでなく、全く関連性を見出せないような図書の推薦も目立つという現象が指摘された。その原因として、原田は、大学図書館ではレポート作成などの一時的な理由で図書が利用される場合があり、そのような図書が影響して利用者の関心にそぐわない図書が推薦されてしまったのではないかと述べ、重みづけなどの工夫によりこれらを貸出記録から取り除くことが必要ではないかと考察している。この研究から、協調フィルタリングの手法を図書館に適応させるためには、図書館特有の貸出傾向を把握した上で、貸出データの取り扱い方を十分工夫しなければならないことがわかる。

2.4 ハイブリッド型フィルタリング

ハイブリッド型フィルタリングは、内容ベースフィルタリングと協調フィルタリング両方式の短所を補いかつ長所を活かす目的で、両方式を結合した方式である¹⁸⁾。組合せの方式は多様に存在するが、推薦アイテムの性質によって両

方式を適切に使い分けることで、内容的関連の高いアイテムと、意外性のあるアイテムをバランスよく推薦できると考えられる。

Zan Huang らは、内容ベースフィルタリングと協調フィルタリングをグラフモデル上で統合したハイブリッド型図書推薦システムを考案している¹⁹⁾。このシステムでは、主にキーワードによる書籍の内容情報や顧客の個人情報をもとに顧客と書籍を特徴ベクトルで表し、顧客間の類似度および書籍間の類似度をそれぞれ計算する。そして、顧客情報・書籍情報をノード、類似度をリンクとして、別々のグラフ上でそれぞれモデル化し、さらに顧客ノードと書籍ノードとの間を購入履歴を表すリンクでつなぎ、2層グラフ上の探索から推薦すべき書籍を選出している。この手法は、図書館で個人化した推薦を行うことのできるハイブリッドシステムの試みとして注目されるが、実験自体がオンライン書店上で行われ、購入履歴を図書館での利用記録と同等のものとしてシステムを構築しているため、図書館特有の利用傾向を十分反映していないという点が惜まれる。

2.5 その他の図書推薦システム

そのほか、図書館の推薦システムとして、一般的な内容ベースフィルタリングや協調フィルタリングを用いない事例もみられる。

ハダースフィールド大学図書館の推薦システム²⁰⁾、およびカリフォルニア大学電子図書館プロジェクトで構築された推薦システム²¹⁾では、ともに貸出記録を用いているが、推薦手法としては、各図書に関して同時に貸し出された図書を共起情報として記録し、共起頻度の高いものから順に、その図書の利用者へ提示するというより単純なアプローチをとっている。そのため、上記のフィルタリング手法のように個々の利用者へ対応した推薦は行えないと考えられるが、利用者個人のプロフィールを作成する必要がないことから、厳密な個人情報を要求せずに推薦を実現できる。なお、カリフォルニア大学の実験では、前出の原田の研究と同じく、学生の専門に関係なく利用が集中するような図書などの影響で、やはり貸出記録と全く関連性のない図

書が推薦されたことが明らかにされている。

カールスルーエ大学図書館の図書推薦システムは、貸出記録ではなく、利用者のOPACの検索ログを収集・活用することにより推薦を実現している²²⁾。このシステムでは、同一の検索セッション内において利用者から同時に2点以上の資料の詳細画面が閲覧されたとき、それらを共起情報として記録し、一つの共起関係マトリクスへと集約する。そして、マトリクスのデータの解析により、関連度の高い資料のペア群を抽出して利用者に推薦していく。また、San-Yih Hwang らは利用者の目録検索ログに加えて、タイトルやキーワードなど資料の内容情報も併用した文献推薦システムを考案している²³⁾。システムは、検索ログと内容情報を併せて参照しながら類似する文献のクラスター群を形成し、現在の利用者の検索行動との類似性から推薦すべき文献を選出している。検索ログは利用者の行動履歴という点では貸出記録と同等のものであると考えられ、Hwang の推薦システム、およびカールスルーエ大学図書館の推薦システムは、貸出記録によらずにレコメンドサービスを行うことの可能性を示しているといえよう。



第3図 カールスルーエ大学図書館の用いる推薦システム (BibTip)

2.6 図書の推薦のためのシステムの改善点

前節に示す例から、図書館において貸出記録を用いて推薦を行うためには、図書の購入とは異なる図書館での利用の特徴に対応する必要があることがわかる。

たとえば、貸出記録には利用者にとってそれほど興味のない図書も含まれるというのは、その問題のひとつである。書店においては、書籍は有料の商品であることから利用者の購買が厳

選され、その購入記録も利用者の嗜好を強く反映している可能性が高い。一方、図書館の場合、貸出行為と貸出図書の有用性とは必ずしもイコールではないため、貸出記録の中に利用者にとって重要なものとそうでないものが混在する可能性があると考えられる。

前節で示す研究の多くでは、図書の貸出行為自体が推薦の基準と捉えられており、貸出記録に含まれる図書自体の差別化が十分行われているとはいえない。しかし上記のような理由から、図書館において有益な推薦を行うためには、貸出記録をそのまま用いるのではなく、図書の重みまで考慮して全ての貸出記録から無用な図書を除外しなければならないだろう。

本研究では、貸出記録から利用者にとって重要性の低い図書を取り除くことに主眼を置き、従来の研究ではほとんど行われていなかった図書データへの重みづけの変更を試みた。

3. 貸出記録を用いた図書の推薦実験

3.1 実験に用いたデータおよびシステム

本実験では、ある私立大学図書館の実際の貸出記録約 90 万件を用いて、協調フィルタリングの手法で図書の推薦を行う。その際、貸出図書の属性、貸出時期によって、利用者と図書とのつながりに重みを設定できるものとした。

本研究で貸出記録を推薦データの決定に用いた大学図書館では、奉仕対象となる学生数が約 3000 名、学生の専攻する分野は人文科学、社会科学、自然科学と多岐に渡っている。蔵書規模は、和洋図書あわせて約 65 万冊、年間入館者数は約 30 万人である。本実験では 1991 年から 2007 年までの期間の貸出記録からリザーブブック、雑誌、また延長手続きによる貸出し、長期貸出、図書以外の貸出などを除外した学部生及び大学院生の貸出記録のうち、在学中の貸出冊数が 200 冊以上 500 冊未満である学生および大学院生の貸出記録、計 272,751 件を使用した。

また、貸出記録をもとにした図書の推薦の判定を行う被験者としては、利用者本人に貸出記録の提供を行うサービスがある大学に在籍する東京都内の大学生で、公開できる貸出図書数が

21 冊以上ある計 10 名とした。被験者自身が貸出記録を入手し、その中から公開することを許諾した図書のリストをもとにして図書の推薦を行った。利用者の専攻分野および過去の貸出件数は第 2 表に示すとおりである。図書推薦のもととなる貸出記録を提供いただいた大学とは別の大学に所属する被験者も含まれている。その被験者については、貸出記録中の図書と ISBN が一致する図書のみを対象とした。

第 2 表 被験者の専攻分野および貸出件数

利用者	専攻分野	貸出件数
I	理工学	67
II	法律学	118
III	英米文学	22
IV	政治学	34
V	政治学	94
VI	英米文学	90
VII	図書館・情報学	28
VIII	図書館・情報学	51
IX	英米文学	116
X	経済学	65

図書の推薦に用いたシステムは、2005 年に Stéphane Droux により開発・公開された、Web サイト上で協調フィルタリングによる推薦を可能にするシステムである Vogoo PHP Pro v2.2²⁴⁾ をもとにし、利用データにあわせたカスタマイズや処理速度の向上などのための修正を行って使用した²⁴⁾。

このシステムでは、利用者と貸出される図書の関係の強さを 0 から 1 までの重みとして付与することができる。図書での利用においては、利用者による図書の興味についての評価とともらえることができる。ただし、一般的な貸出記録においては、重みとして使うことができるような、図書に対する直接的な評価値は存在しない。本実験では、このような重みについて、「図書の分野」、「時間 (貸出日付)」という属性を設定して実験を行った。

3.2 貸出データへの重みづけ

「図書の分野」については、学生が、専攻によって集中して利用する図書の分野（NDCの10区分を用いる）が異なるという星野らの分析結果¹⁰をもとに重みを設定した。すなわち、貸出記録に占める分野的比率の高い図書ほど、学生にとって関心の高い図書であると考え、重みを大きくした。

「時間（貸出日付）」については、学生の興味が学年や時期によって変化することに着目して重みを設定した。たとえば、衛藤らは、学生は学年が上がるにつれより専門的な内容の図書を求めるようになることに言及しており¹¹、また、原田の推薦実験から、学生が一時的な目的で図書を利用する可能性があることも明らかとなっている¹²。このような傾向を踏まえ、被験者が最近利用している図書には興味を持っている可能性が高いことから、貸出日付のより新しい図書ほど重みを大きくするものとした。

また、レポートの執筆などのための利用を中心に、学生の興味は貸出時期によって大きく変化することが予想される。その場合、同じ学生であっても貸出時期が異なれば、別の利用者にとらえることができるほどの相違とも考えられる。そこで、「時間（貸出日付）」を基準として重みを付与する手法として、学生の興味が大きく変化した場合には別の学生の利用とみなすものとし、各利用者の貸出記録を分割し、利用時期ごとに別の利用者番号を付与して同様の実験も行った。

この「図書の分野」と「時間」に着目した貸出記録の取り扱いの組み合わせとして、以下に示す5つのパターンを設定して推薦実験を行った。

1) パターンO

先行研究と同じく貸出データ全てに一律の重み1.0を与える方法で、重みを変更する他のパターンによる推薦との比較のために行う。

2) パターンA

「図書の分野」を考慮した重みづけ。各利用者の貸出記録について、最も多く借りられた分

野の図書には1.0、次に多いものには0.5、三番目に多いものには0.25...というような重みを与える。

3) パターンB

「時間」を考慮した重み付け。各利用者の貸出記録をひと月単位で分割し、別の利用者番号を付与することで、別々の貸出記録として扱う。たとえば、1年間で12カ月全ての月に図書を利用した学生の貸出記録は、12の別々の貸出記録へと分けられる。このパターンでは貸出記録を分割することによる効果を検証するために、貸出記録の重みは全て1.0とした。

4) パターンC

「時間」を考慮し、パターンBと同じく貸出記録をひと月単位で分割するが、同時に被験者が最近借りた図書ほど重みを大きくするという重み付け。すなわち、興味の変化を考慮するとともに、被験者が過去に借りた図書は興味を失ったものとみなし重視しないというものである。具体的には、まず、被験者の記録のうち貸出日付に着目し、一番最近図書を借りた月の図書に対して1.0の重みを付与し、二番目に新しい貸出月に含まれる図書に0.5、三番目以降の貸出月の図書には0.25, 0.125...のように漸次重みを小さくして与えた。なお、貸出記録中の図書の重みは全て1.0とした。

5) パターンD

「図書の分野」と「時間」の両方を踏まえた重みづけ。すなわち、貸出記録はひと月単位で分割し、被験者の貸出記録については、パターンAと同様に貸出の多い分野の図書ほど大きな重みを与える。なお、貸出記録中の図書の重みは全て1.0とした。

各パターンから選出した推薦図書上位20冊ずつを各被験者へ提示し、評価の高い方から順に「強い興味」、「読みたい」、「興味なし」の3件法による評価を求めた。利用者に評価を求め際には、図書の情報として、「タイトル」「著者名」「出版者名」「出版年」の4項目を提示した。内容の判断ができないなどの理由で評価できなかった図書は「評価不能」とした。

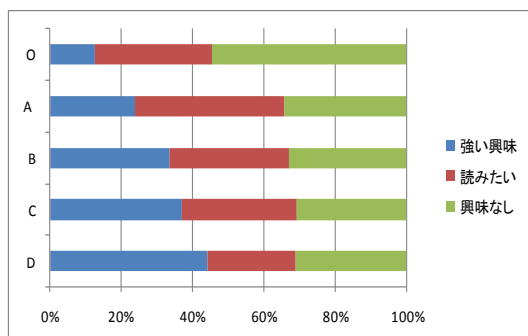
また、利用者にはレコメンドサービスや推薦結果に関して評価実験後にインタビューを行った。インタビューと推薦図書への評価、および推薦図書と貸出記録との関連性といった観点から、推薦パターンの有効性を検証した。

3.3 推薦結果と利用者の評価

実験パターンごとに得られた推薦図書に対する被験者全体の評価結果を第3表および第4図に示す。

第3表 各パターンの推薦結果の利用者の評価

Pat	強い興味	読みたい	興味なし	評価不能
O	17(12.7%)	44(32.8%)	73(54.5%)	66
A	47(24.0%)	82(41.8%)	67(34.2%)	3
B	67(33.5%)	67(33.5%)	66(33.0%)	0
C	73(36.9%)	64(32.3%)	61(30.8%)	2
D	87(44.2%)	49(24.9%)	61(31.0%)	3



第4図 各パターンの推薦図書に対する利用者全体の評価 (評価可能図書についての比率)

第3票および第4図に示すように、「強い興味」、「読みたい」を好意的な評価であると考え、全ての図書に同じ重みづけを行ったパターンOを除くすべてのパターンにおいて、評価可能な図書のうちの65%以上が好意的な評価を得られた。パターンOの推薦図書に対する好意的評価が平均45.5%にとどまったことと比較すれば、図書データの重みづけなどを工夫することによって、利用者のニーズにより適合した推薦を行えるということが示されたといえる。特に、貸出記録を月単位で分割して使用したパ

ターンB, C, Dは、好意的な評価を受けた推薦図書の割合がいずれもパターンAよりさらに1.2~3.4ポイント高くなっていた。

なお、パターンOの結果は、第2章に示す原田が約90万冊の図書をもとに行った同様の実験結果¹⁷⁾と比較して、評価不能と判定された図書の数が多く、また、「強い興味」「読みたい」と判定された図書の比率が低い結果となった。このことは、貸出記録を用いた図書の推薦を行う際には推薦結果の判定のもととなる貸出記録の数が多くの方が有効であることを示しているともいえよう。

3.4 分野別の図書の推薦結果

パターンA, およびパターンDの場合の各被験者の分野別の貸出冊数と推薦結果との対応を、それぞれ第4表および第5表に示す。

第4表に示すように、パターンAの推薦結果では、被験者が利用した図書の分野が集中しているほど重みを高く設定したが、推薦される図書の分野は、被験者によってはそれほど集中しない傾向が見られた。特に、専攻分野が社会科学系である被験者(II, IV, V, X)について、貸出内容に一致した3類図書が多く推薦されたが、その他の被験者に対しては、利用した図書の分野よりも1類と3類の図書が多く推薦される傾向が見られた。これは、多くの学生が履修する科目のレポートなどで、同じ図書が多くの人々に利用されたため、どのような図書を利用している人にとっても同じ図書が推薦される結果となってしまったと考えられる。

一方、貸出記録を分割して用いたパターンDでは、利用者平均で4.3分野にわたる図書が推薦され、幅広い分野にわたる推薦結果がみられた。また、パターンAと比較してパターンDの方が高い評価が得られた。このように、貸出記録を分割して用いることは、一時的な目的や関心から利用された図書を取り除くだけでなく、一部の利用者の貸出による偏りを少なくするためにも非常に有効であるということが考えられる。そしてその結果として、推薦に不要な図書を減らし、被験者の評価向上につながったものと考えられよう。

第4表 利用者の分野別貸出冊数とパターンAによる分野別推薦図書

利用者		0類	1類	2類	3類	4類	5類	6類	7類	8類	9類	不明	分野数	計
I	貸出	23			1		26		7	10			1	67
	推薦結果		19											1分野
II	貸出	25	1	5	48	4	15	6	6	6	2			118
	推薦結果		10	1	9									3分野
III	貸出										22			22
	推薦結果		20											1分野
IV	貸出	9		2	18			3	1		1			34
	推薦結果		1		19									2分野
V	貸出	5	4		67	3	8	5		2				94
	推薦結果		8	1	11									3分野
VI	貸出	1	3	2	3				6	2	73		2	90
	推薦結果		16	1	1									3分野
VII	貸出	8	4	3	3				6	4				28
	推薦結果		1	1	13				4		1			5分野
VIII	貸出	24	1		9		3	6	2	4	2			51
	推薦結果		8	5	7									3分野
IX	貸出	12	4	3	9	1	1		4	1	81			116
	推薦結果		18	2										2分野
X	貸出	7		2	41	2	7	6						65
	推薦結果		2		18									2分野

第5表 利用者の分野別貸出冊数とパターンDによる分野別推薦図書

利用者		0類	1類	2類	3類	4類	5類	6類	7類	8類	9類	不明	分野数	計
I	貸出	23			1		26		7	10				67
	推薦結果	1				5	9		2			3		4分野
II	貸出	25	1	5	48	4	15	6	6	6	2			118
	推薦結果		2	1	16						1			4分野
III	貸出										22			22
	推薦結果	2	1	3							14			4分野
IV	貸出	9		2	18			3	1		1			34
	推薦結果	5	1		13			1						4分野
V	貸出	5	4		67	3	8	5		2				94
	推薦結果		4	2	14									3分野
VI	貸出	1	3	2	3				6	2	73			90
	推薦結果	1	3		5				4		7			5分野
VII	貸出	8	4	3	3				6	4				28
	推薦結果	14	1		3	1			1					5分野
VIII	貸出	24	1		9		3	6	2	4	2			51
	推薦結果	8	2	3	2				4		1			6分野
IX	貸出	12	4	3	9	1	1		4	1	81			116
	推薦結果		2		1				2	1	14			5分野
X	貸出	7		2	41	2	7	6						65
	推薦結果	1		2	16			1						4分野

3.5 レコメンドに関する利用者の評価

さらに、各パターンの推薦結果を見た上で、被験者に「好みにかかわらず自分の貸出記録と特に関連があると考えられたパターン」、および「貸出記録との関連性の高さにかかわらず自分が興味深いと思えたパターン」をそれぞれ2つずつ回答してもらった。

その回答の対応を第6表に示す。表にみられるように、全ての被験者が、関連性の高さ興味の一貫したパターンを1つは挙げており、2つとも一致していた被験者は半数を占めた。さらに、パターンC、Dのいずれかまたは両方について興味があるという被験者が多かった。具体的には、パターンCを関連性があると判断した被験者が6名、そのうち5名がこのパターンに興味があると回答し、パターンDを関連性があると判断した被験者は9名、そのうち6名がこのパターンに興味があると回答した。

第6表 利用者の選択した、貸出と関連性の強いパターンおよび興味深い推薦パターン

利用者	関連がある					興味がある				
	O	A	B	C	D	O	A	B	C	D
I			○		○			○		○
II				○	○		○		○	
III			○		○				○	○
IV				○	○			○	○	
V				○	○		○		○	
VI			○		○			○		○
VII	○				○	○				○
VIII				○	○				○	○
IX				○	○		○			○
X		○		○			○		○	

貸出記録を同じように分割した手法でありながらも、パターンBによる推薦は、関連性や興味の高さに関してパターンC、Dよりも評価が低かった。この理由としては、パターンBでは貸出時期が異なるが利用者が共通であるという図書が推薦される可能性が高く、その結果と

して、利用者の関心をひく図書とそうでない図書が混在してしまったと考えられる。

さらに、パターンCおよびパターンDのいずれかのみを評価した被験者に対して理由を聞いた。その結果、パターンCのみを評価する被験者からは、過去の貸出に関連があるものの、レポート作成などのやむを得ない事情から利用した図書に関する推薦が多かったため興味を持てなかったとする意見が挙げられた。パターンCは本実験において、好みにかかわらず最近の貸出図書との関連が強い推薦が多かったという意見が多数聞かれており、このパターンに興味を示した被験者の多くは自身の最近の興味に適した推薦を望んでいた。

一方、パターンDのみを評価する被験者からは、継続的な興味関心に沿って一定の分野の図書をこれまで利用してきており、それに関連する推薦の多いパターンDに必然的に良い評価を与えている場合が多かった。

このように、図書館を利用する利用者には、
1) 興味の対象が変化しており、過去よりも最近の図書館利用を反映した推薦を好む利用者と、
2) 長期間にわたって一貫した興味関心から図書館を利用している利用者、の両者が存在すると考えられる。それぞれがパターンCおよびパターンDの推薦への評価につながっている。さらに、パターンDを好意的に評価する被験者であっても、現在の興味に近い図書を推薦してほしいという意見を持つ人も存在していた。このように、図書の推薦の際には、ひとつのパターンだけでの推薦ではなく、いくつかの選択肢を用意することの必要性も示していると思われる。

3.6 レコメンドサービスそのものに関するインタビュー調査

図書館においてAmazonのようなレコメンドサービスを行うことへの賛否について尋ねたところ、多数の利用者からレコメンドサービスの導入に賛成する、または興味があるという肯定的な意見を得られた。その理由としては、効率的によく図書を探ることができる点や、自分が探している図書に関連した推薦があると嬉しいという点のほか、興味のあるテーマについて違

った切り口から図書を探することができる点、興味深い図書を発見するきっかけが得られる点も挙げられた。つまり、レコメンドサービスを実現することによって、利用者の求める図書を積極的に提供できるとともに、ときに資料のブラウジングのような効果を発揮することも期待できよう。

逆に、レコメンドサービスは図書館をあまり利用しない人にとって煩わしいのではないかという消極的な意見や、必ずしも必要というわけではないという中立的な意見なども存在した。このような利用者の意見も尊重し、事前に利用者に許可を求めたり、強制的なサービスとしないといった配慮をすべきであろう。

そのほかの意見としては、レコメンドサービスに対する要望として、本実験で提示した推薦図書の情報だけでなく、図書の表紙画像、著者のプロフィール、あらすじや要約、利用者からのレビューがあれば評価の参考になったという意見も挙げられた。確かに、利用者が推薦図書に対して適切な判断を下すためには、本実験で提示した4項目だけでは不十分であり、利用者に親切なレコメンドサービスを行うためには、このような情報を提供することも不可欠となってくるだろう。

さらに、推薦図書のリストを、テーマや出版年や専攻との関連性などによってソートできるようにしてほしいという意見も挙げられた。専攻との関連性によるソートを実現するためには、何を基準にして優先する推薦図書を決定するか検討していかなければならないが、テーマや出版年などによるソートは現状でも十分可能であろう。本実験では利用者に提示した図書以外にも多くの推薦図書が選出されており、ソート機能による推薦結果の調整により、当初推薦されなかった図書から適切な図書を発見することも考えられる。

また、本実験の推薦結果は娯楽性を持つ推薦としては評価できるが、本格的な学問や研究に用いる推薦としては不十分であるという指摘をする被験者も存在した。このようなニーズに対して有効な図書の推薦を行うためには、推薦す

る図書をより厳選するような重みづけなどの工夫や、内容ベースフィルタリングを用いて新しい図書を推薦できるようにすることも検討していく必要があるだろう。

4. 貸出記録を用いた推薦システムの可能性

本研究では、協調フィルタリングと貸出記録を用いながら大学図書館での図書の推薦実験を行った。特に、利用者にとって不要な図書が推薦されることを防ぐ目的で、貸出記録に含まれるノイズの原因となる図書を取り除くことに焦点を当てた。そのために利用者の図書館利用動向に着目し、利用される図書の分野の比率や貸出日付の新しさによって貸出データの重みづけを行い、貸出データの間に差別化を図った。さらに、利用者の興味の変化を考慮した手法として、同一人の貸出記録を分割して用いることも提案した。

実験の結果、新しい手法による推薦図書は、重みが全て一律である貸出記録を用いた推薦図書よりも利用者から総じて高い評価が得られ、貸出記録の取り扱いを工夫することで不要な図書を除き、利用者にとってより適切な推薦を行えることが明らかとなった。特に、貸出記録を分割することは、一部の利用者の貸出記録に推薦結果が極端に影響されることを防ぐためにも有効であることが考えられた。そして、図書の重みづけと分割手法を組み合わせることで、分野を重視した推薦や時間軸を重視した推薦のような、特定の意図に沿った図書の推薦が可能であることも示すことができた。

ただし、利用者からの評価やインタビュー回答から、推薦図書の中には未だに利用者の興味をひかない図書が含まれている点や、研究に直接役立つ図書があまり推薦されていなかったという点が指摘されていた。貸出記録を用いてさらに有効な図書の推薦を行っていくためには、本実験よりも一層綿密に貸出データの取り扱いについて検討していく必要があるだろう。たとえば、分割した貸出記録にもさらに重みづけを行うことや、時期によって分割する範囲を変えるといった手法などが考えられよう。

また、本研究では不要な図書を除外することに焦点を当てたが、今後は利用者によって異なる情報ニーズを考慮したシステムを検討していくことも重要である。本実験においても推薦図書に対して様々なニーズや好みが見られたが、利用者の希望をどこまで尊重するかによって、貸出データの取り扱い方は変わってくる。多様な目的を持つ図書館利用者に対して柔軟にレコメンドサービスを行っていくためにも、専攻や図書館利用目的などによって推薦対象を変化させることや、推薦結果に絞り込みやソートをかかけるといった機能の充実が求められよう。

Amazon や Google のような利用者志向のサービスが当たり前となる流れを受けて、OPAC もようやく他の Web サービスのように便利で親しみやすい機能やデザインを実現しつつある。レコメンドサービスはそのようなサービスの一つとして、図書館へのナビゲートとして有効に働くであろうし、利用者を図書館の未知の領域へ導くことのできる手段ともなり得るだろう。

本研究は、特に貸出記録の取り扱いという点から、図書館にふさわしい推薦システムの可能性を示すことができたと確信している。

5. 謝辞

本研究の実験に際して貸出記録の利用をご許可いただいた大学図書館の関係者諸氏、被験者として貸出記録を提供いただき、評価に協力して下さった方々に深く感謝いたします。

6. 注・引用文献

- 1) 日本情報処理開発協会. 情報化白書 2009. 東京, 日本情報処理開発協会, 2009, 353p.
- 2) 久保山健. 次世代 OPAC を巡る動向: その機能と日本での展開. 情報の科学と技術. vol.58, no.12, 2008, p.602-609.
- 3) 林賢紀. 次世代の図書館システムと OPAC. 医学図書館. vol.55, no.1, 2008, p.121-126.
- 4) 岡本真. 利用者の目からみた図書館の目録——評価する点,改善すべき点,期待する点. 現代の図書館. vol.41, no.4, 2003, p.217-221.
- 5) 岡本真. Web2.0 時代の図書館—

Blog.RSS.SNS.CGM. 情報の科学と技術. vol.56, no.11, 2006, p.502-508.

- 6) 出版月報 1 月号. 東京, 出版科学研究所, 2006. 月刊
- 7) 日本図書館協会. 日本の図書館統計. [2007-12-12], <<http://www.jla.or.jp/statistics/>>
- 8) 渡邊斉志. 知的自由の陥穽: 利用情報保護思想が公立図書館に及ぼす影響の分析. Library and information science. no.58, 2007, p.103-115.
- 9) 安東奈穂子, 池田大輔, 田中省作. 電子図書館と利用者のプライバシー: 履歴・個人情報の保護と利用の両立を目指して. デジタル図書館. no.30, 2006, p.62-71
- 10) 星野雅英, 渡邊真由美, 風巻利夫, 原香寿子. 図書館の統計 東京大学総合図書館における入館・貸出統計データ分析の試み—中央図書館としての役割を考えるために. 大学図書館研究. no.82, 2008, p.1-11.
- 11) 衛藤広隆, 吉田美紀子. 大学図書館と貸出—中規模私立大学図書館の立場からの分析. 私立大学図書館協会会報. no.96, 1991, p.114-124.
- 12) 佐浦敬之. 貸出履歴の活用に関する意識について. 図書館の自由 別冊. [2009-10-12], <<http://www.jla.or.jp/jiyu/taikai2008kiroku.pdf>>
- 13) 徳永健伸. 情報検索と言語処理. 東京, 東京大学出版会, 1999, 234p.
- 14) 寺野隆雄. 情報推薦システム. e ビジネスの理論と応用. 東京, 東京電気大学出版局, 2003, p.59-87.
- 15) 廣岡康雄, 寺野隆雄, 大塚雄吉. 意外性の高い情報を提供するレコメンダー・システム. 知識ベースシステム研究会. no.47, 2000, p.61-66.
- 16) Greg Linden. ; Brent Smith. ; Jeremy York. Amazon.com Recommendations : Item-to-Item Collaborative Filtering. February 2003 issue of IEEE Internet Computing. 2003, p.76-79.

- 17) 原田隆史.図書館の貸出履歴を用いた図書の推薦システム.デジタル図書館. no.36, 2009, p.22-31.
- 18) 石川徹也, 宇田隆幸.情報フィルタリングの利用システム: 情報推薦システム. 情報の科学と技術. vol.56, No.10, 2006, p.458-462.
- 19) Z Huang. ; W Chung. ; T Ong. ; H Chen . A graph-based recommender system for digital library. ACM/IEEE-CS Joint Conference on Digital Libraries. vol.JCDL2002, 2002, p.65-73.
- 20) Dave Pattern. using "circ_tran"to show borrowing suggestions in HIP - "Self-plagiarism is style".[2009-11-10], <<http://www.daveyp.com/blog/archives/49>>.
- 21) Collen Whitney.;Lisa Schiff.The Melvyl Recommender Project: developopting library recommendation services. D-Lib Magagine.vol.12, no.12,2006. [2009-11-04], <<http://www.dlib.org/dlib/december06/whitney/12whitney.html>>
- 22) Michael Mönnich. ; Marcus Spiering. Adding value to the library catalog by implementing a recommendation system.D-Lib Magagine.vol.14, no, 5/6, 2008. [2009-11-04], <<http://www.dlib.org/dlib/may08/monnich/05monnich.html>>
- 23) San-Yih Hwang ; Shi-Min Chuang. Combining article content and Web usage for literature recommendation in digital libraries. Online Information Review.vol.28, no.4, 2004, p.260-272. [2009-7-13], <<http://www.emeraldinsight.com.kras1.lib.keio.ac.jp:2048/Insight/viewPDF.jsp?contentType=Article&Filename=html/Output/Published/EmeraldFullTextArticle/Pdf/2640280402.pdf>>
- 24) Stéphane Droux.Vogoo -Recommendation Engine &Collaborative Filtering. [2009-12-18], <<http://www.vogoo-api.com/index.php>>