

図書館貸出データを用いた学習グループ候補の発見 — 利用者の社会的ネットワーク発見への試み —

Discovery of Study Group Candidates from Library's Loan Records
— An Attempt for Finding Social Networks of Library Patrons —

南俊朗, 馬場謙介

Toshiro Minami and Kensuke Baba

【要 約】

図書館が学習グループの構成支援サービスを提供することにより図書館は利用者にとっての“ソーシャルメディア”になることができる。これはネットワーク時代の図書館における新規サービスへの1つの試みとしての意義があり、また近年の大学図書館において利用者が長時間図書館に滞在するような環境づくりの切り札的存在となっているラーニングコモンズ (LC) 空間の有効活用にも寄与することになる。このような背景の下、本稿は図書館の貸出記録に基づいた学習グループの構成という手法について議論し、1つの方法を提案する。本手法においては、学習グループのメンバは貸出記録から定義される利用者の興味分野の類似性や専門度等に基づいて選定される。このような目的やアプローチによる研究はこれまで十分なされておらず、本論文の議論もまだ初歩的レベルにはあるものの、将来の図書館サービスとして大きな潜在的重要性をもつ。

キーワード： 図書館データ解析, 利用者/学生のプロフィール, 興味分野, 専門度, 社会的ネットワーク発見

[Abstract]

Library's study group formation service can make the library work as a "social media" for patrons, which is an attempt to develop a new library service in the network age. It can be considered also to contribute to well-utilizing the learning commons (LC) spaces of the library. LC is a trend for libraries for attracting the patrons and encourage them to stay in the library for a long time. Based on these background motivations, we discuss and propose a method of study group formation by libraries, based on the loan records. In the proposed method, study group members are chosen based on the similarity of patron's interest area profile, expertise level, etc., which are definable using the loan records. Even though the research on this approach has just started and thus is still in the very early stage, we are convinced of its great potential for valuable library services in the future.

Keywords: Library Data Analysis, Patron/Student Profiling, Interest Area, Expertise Level, Social Network Discovery

1 はじめに

学習グループの発見という本研究には2つの背景がある([12]も参照)。第1の背景は図書館の社会的役割の変化に関してである。ICT(Information and Communication Technology)の発展と普及に伴い、我々の社会は大きく変化しつつある。それに伴い図書館の社会的役割は変化しつつある。図書館学の第5法則[13]にもあるように「図書館は成長する生命体」であり、利用者の変化に適應する努力を続けている。

ネットワーク社会で注目を集めているトピックとして“ソーシャルメディア(Social Media)[2]”がある。現在ではスマートフォンやPDA(Personal Digital Agent)機器、タブレットなどのモバイル情報機器を日常的に持ち歩くことが普通のこととなった。その結果我々はネット接続を行いつつどこからでも情報サービスにアクセスすることができる。そのようなユビキタスな環境はICT技術を背景としてソーシャルメディアサービスにより我々がリアルタイムでコミュニケーションを取ることを後押ししている。

ソーシャルメディアのオンラインでの有用性を考えると、それはオフラインでも有益であるものと推定できる。歴史的に図書館は図書や雑誌などの情報資源を利用者に提供することにより利用者の情報や知識に対する需要を満足させる社会的役割を果たしてきた。したがって、潜在的ネットワーク関係に基づき、利用者同士をつなげるサービスを提供し、利用者が協力して学習したり、意見交換したり、共同活動を行ったりすることは、ネット時代における図書館サービスの自然な拡張である。これが第1の背景である。

ICTの発展は他方図書館利用の減少の一因にもなっている。以前、我々は必要な情報を図書や雑誌、その他の資料から入手するために図書館を訪問した。それとは対照的に現在は、ポケットからスマートフォンなどのモバイル情報機器を取りだし、検索エンジンに問い合わせるだけで必要な情報を入手できるようになり、図書館を訪問する必要性はなくなった。多くの人々にとって図書館は最早必要不可欠なものではない。

このような環境の変化を考慮し、図書館は利用

者にとって訪問し、滞在する価値のある場所として認識してもらうための努力を行ってきた。その一環として、特に大学図書館を中心に Learning Commons(LC)と呼ばれる特別なエリアを設定することが最近増えつつある。そこでは利用者は自由にテーブルの配置を変えたり、本を音読したり、ゼミや学習会などを開いて議論をしたりして、ある程度の音や声を出すことが許される。今では大学図書館は新築や改築の際LCエリアを設置することが多い。LCは学生の能動的な学習(Active Learning)を支援する有効な施設として、大学図書館にとって必要なものであると見なされている。これが第2の背景である。

これらの背景となる動機に基づき、本稿では、学生の学習グループの構成や支援を行う図書館サービスに注目する。本サービスは図書館にとって、新しい利用者サービスを提供する意義だけではなく、LCエリアを有効活用するためにも有益である。そのための手段として、全ての図書館が入手可能である貸出データを活用し学習グループの候補を生成・提案する手法を追究する。本稿の手法では、貸出データを用いて定義することのできる利用者の興味分野や専門レベルなどの類似度に基づいて1つのグループを構成するのにふさわしいと考えられるメンバを選択し、1つのグループとして提案する。

図書館の学習グループ形成サービスは利用者側と図書館側の双方にとって利点がある。特に大学図書館の、利用者にとっての第1の利点は、学習グループとして適切な仲間を探すことが容易になることである。もし学生が学習グループの仲間となる他のメンバを自ら探そうとしても、時間や手間がかなりかかってしまうであろう。図書館が、保持するデータを基に、その利用者の中から適切であると考えられる他メンバを選択し、推薦するならば、推薦される仲間の学生は、同じ学習グループの仲間として、学習への興味や、分野、知識レベルなどにおいて、ある程度保証されることが期待できる。そのような、いわば質の揃った学習仲間を自らの努力だけで探し出すことは極めて面倒である。

第2は、図書館が推薦する学習グループのメン

は学習の仲間として十分信頼に足る学生が多いものと考えられることである。というのは推薦されるメンバは図書館の利用者の中から選ばれるため、一般の学生の中から選ばれる場合と比べて、学習に熱心であり、また、真面目な性格の学生であることが期待できる。図書館による学習グループメンバの推薦は、そういった信頼感がある。

第3は、図書館の責任で学習グループを推薦するため、学習に参加する学生メンバは、図書館からのさまざまなサポートが期待できることである。たとえば、学習に適した学習室を確保してくれるであろうし、また、図書館の司書（特にサブジェクトライブラリアンと呼ばれる、その分野を専門とする司書）からのアドバイスや学習の補助があるかも知れない。このような支援が期待できることは、図書館による学習グループ候補推定の大きな魅力となる。

更に、学習グループのメンバは全ての図書館利用者の中から選ばれるため、同じ学習グループメンバの中には通常の状態では決して出会うことがなさそうな学生が含まれているかも知れない。そのように通常の状態では知り合うことのないような良い仲間を推薦してもらえることは、図書館による学習グループの推薦サービスのもう1つの可能性である。

一方、学生のための学習グループ形成と推薦サービスは図書館にとっても大きな利点がある。既に指摘したように、図書館はこれまで利用者（たとえば学生）が図書館に魅力を感じ、来館し、学習や研究、読書、その他の活動のために図書館に長時間滞在できるような滞在型図書館としての環境整備に大きな努力を傾けてきた。学習グループを形成し、学習グループによる共同学習を支援することは、多くの学生が図書館を訪問し、学習室を利用したりすることにより、このような努力を更に後押しすることに繋がる。

特に、大学やその他の教育機関の図書館の場合、近年、学生達をより「能動的学習」を実践するための環境を整備することが求められている。その観点から見ると、図書館による学習グループの形成サービスは、能動的学習を促すための環境作りの1つであると見なすことができ、図書館の責務

を果たすことに貢献している。それは教育機関である大学の1つの部署として、組織の学習支援能力の向上に役立つことができる。

同様に、学習グループを形成するサービスは利用者にとって最も好まれる図書館サービスとなり得る。それは図書館の利用者（顧客）から更なる信頼を確保するのに役立つものと考えられる。

従来のサービスと比べて一層進歩的な図書館サービスとして図書推薦サービスに関する研究は多数存在する [1][3][15]。貸出データは利用者と図書がどのように関連しているかに関する1次的情報を与えるものであるため、利用者への図書推薦の原データとして非常に利用価値が高い。たとえば、協調フィルタリング (Collaborative Filtering) の手法 [14] を用いることで、利用者に関連していると考えられる図書の推薦に用いることができる。

図書館における学習グループ推薦に関する本稿のようなアプローチの研究は著者達の知る限り未だ他には存在せず、新領域の研究である。本稿のようなアプローチにより一緒に学習を行っていく1つのグループとして本当に適切なグループを見出すことができるのかどうか、図書館が入手可能なデータだけで判断できるものなのか、など多くの疑問点が残る。しかし、参加者の興味分野がばらばらなグループ分けを行った場合と比較して、興味分野やレベルなどの一致度が高いグループ分けの方が学習グループとして成功する可能性は高いであろうと考えることは妥当であろう。本稿は、このような仮定の下、データに基づいた学習グループ候補の選出という手法の可能性を探る第一歩としての意義があるものと考えられる。図書館が学習グループの構成に一役買い、学生達の学習を支援するという新しいサービスの試みは、その潜在的な重要性を考えると、将来的には価値の高い図書館サービスへと発展することが期待できる。

本稿の目的を達成するため、以下のように構成される：まず、第2節においては、貸出データに基づいて定義される利用者の興味分野なる概念を用いて利用者プロフィールを定め、それに関する基礎的な分析を行う。なお、本稿で用いられる貸出データは九州大学附属図書館の中央図書館か

ら提供されたものである。また、2つの利用者プロフィールの類似性を図る指標 [7][8] を定義する。この指標は学習グループのメンバの間の興味分野の類似性の判断に用いられる。

次に、第3節において、利用者の専門性のレベルを図る指標として専門度なる概念を定義する。ここでも貸出データが用いられる。

これらの節における議論を踏まえ、第4節では、可能な学習グループの候補を見つける手法を提案する。まず、興味分野のプロファイルを用いて、プロフィールが類似した利用者を1つのグループとしてまとめることにより学習グループを1次的に選択する。次に、更に詳しく興味分野を分析したり、利用者の専門度をも考慮に入れることで、更に詳細化された学習グループ候補案を作成する。

最後に、第5節において、本稿の議論を総括し、また、将来への方向性を示す。

2 興味分野のプロファイルとその類似性の指標

学習グループのメンバの興味分野が一致もしくは類似しているか否かは学習グループとして上手く働くかどうか大きく影響するものと考えられる。もしメンバの興味分野が大きく異なるならば、一緒に学ぶ内容を選択することが困難であろうし、学ぶ意欲の弱いメンバがいると、他のメンバの学ぶ意欲も殺される恐れがある。したがって学習グループメンバを選択するための第1の条件としてメンバの興味分野が何らかの共通性をもっていることを設定する。

ここで問題となるのは利用者の興味分野をどう定義するか、そしてその類似性の指標を如何に定義するかである。我々は、図書館の貸出データを用い、それを基に利用者の興味分野を把握するというアプローチを採用する。本節では、まず、第2.1節で、貸出データの概要を説明する。次に、第2.2節で、利用者の興味範囲なる概念を定義する。そして、何人かの学生を例に、興味範囲の実例を示し、また、それらを比較する。最後に第2.3節で、興味範囲の類似性の指標を定義し、それに基づいて、学生相互の類似性に関する分析例を示す。

2.1 貸出データの概要

本稿で用いられる図書館の貸出データは2007年度の九州大学附属図書館(中央図書館)のもの([5]-[11])である。1件のレコードは図書(書誌)ID, 図書の分類番号(NDC, 日本十進分類), 呼出番号, 貸出者ID(プライバシー保護のため付け替えられた番号を使用), 所属情報, 貸出者タイプ(学生, 教員, 職員など), 貸出日時, 返却日時などから構成されている。前処理として本論文の処理に必要とされる情報の欠損や不備のあるレコードを削除した結果, 53,182件のレコードが残された。これらのレコードによると, 図書貸出を行った利用者は5,718名である。

貸出を行った利用者を10種類のタイプに分類する: 学部学生(B1からB6), 修士学生(M), 博士学生(D), 教員(P, Professor), その他(O, Other)。貸出レコードに占める貸出者のタイプを調べると, 約半分53%が学部学生による貸出であり, 約1/4である23%は修士学生, そして13%は博士学生によるものであった。すなわち89%もの貸出は学生によるものであり, 大学図書館にとって学生が最大の利用者(すなわち, 顧客)であることを裏付けることができる。この事実も, 学習グループの推薦という図書館サービスが主たる利用者である学生向けのサービスとして妥当であることを示している。

2.2 利用者プロフィール作成のための興味分野

本稿では学習グループのメンバ間に興味分野の類似性を仮定する。類似性を図るための指標を定義する準備として, まず興味分野のプロファイルなる概念を定義し, その後プロフィールの類似性指標を定めることにする。本稿でのアプローチは, 図書館の貸出データを用いて興味分野を推定するものである。それを用いて利用者のプロフィールをモデル化する[7]。「興味分野」を扱うためには, まず「分野」なる概念を定める必要がある。本稿では日本十進分類(NDC, Nippon Decimal Classification)の分類番号を利用することにする。NDCはDDC(Dewey Decimal Classification, Dewey十進分類

法)の日本版であり、日本では多くの図書館で採用されている。したがって、NDCは図書館の全ての蔵書に書誌情報の1つとして与えられており、蔵書内容の分野情報を与えるものとして最適である。

NDCシステムは階層構造になっている。その最上位は000から900までの10のカテゴリからなり、それぞれのカテゴリは更に下位レベルに細分化される。このような細分化が何段階か繰り返される。最上位カテゴリは以下のようにになっている：000 (総記), 100 (哲学・宗教), 200 (歴史・地理), 300 (社会科学), 400 (自然科学), 500 (技術), 600 (産業), 700 (芸術), 800 (言語), 900 (文学)。たとえば、400 (自然科学) のカテゴリの次レベルは次のサブカテゴリに細分化されている：400 (自然科学), 410 (数学), 420 (物理), 430 (化学), 440 (天文学), 450 (地球科学), 460 (生物化学), 470 (植物学), 480 (動物学), 490 (医学)。

このNDCカテゴリを用いて、利用者の興味分野プロフィールをそれぞれのカテゴリに対応する10次元ベクトルとして定める。それぞれの要素は、利用者が借りた図書の中で対応するNDC最上位カテゴリに対応するものの割合とする。形式的に表現すると、利用者 p のプロファイルは次の式により定められる。

$$Prof(p) = \langle r_0, r_1, \dots, r_9 \rangle$$

ここに、それぞれの $i = 0, 1, \dots, 9$ に関して、 $r_i = f_i / \sum_{j=0}^{900} f_j$, $f_i = \#\{l | l : 1 \text{つの貸出記録, } borrower(l) = p, category(l) = i00\}$ とする。ここで $\#$ は集合の要素数を示し、また $borrower, category$ はそれぞれ、貸出者、NDCカテゴリを表す。

したがって、たとえば、ある利用者がNDCカテゴリが000から099.99の図書を10冊、100から199.99の図書を11冊借り、200から299.99の図書を12冊、以下同様に900から999.99までの図書を19冊借りたとすると、そのプロファイルは $\langle 10/145, 11/145, 12/145, 13/145, 14/145, 15/145, 16/145, 17/145, 18/145, 19/145 \rangle$ となる。この定義は利用者グループにも自然に拡張可能である。すなわち、上記定義を、「利用者が借りた図書」の部分で「当該利用者グループに属するある

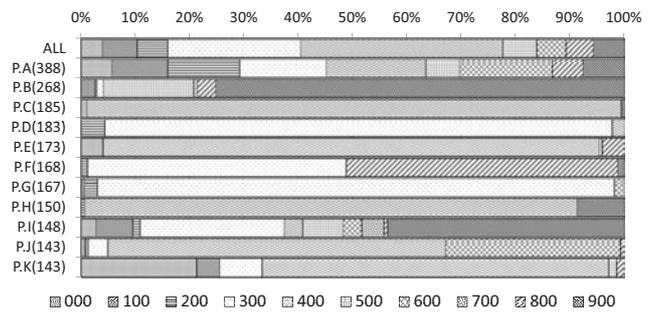


図 1: 興味分野のプロファイル (全ての利用者学生および貸出冊数順でのトップ 11 名の学生)

利用者が借りた図書」と置き換えることで得られる定義である。このような定義の拡張は、利用者プロフィールの類似度や、利用者の知識レベル、学習能力、学習スタイルなど他の概念に関しても同様に可能である。

図 1 に貸出数トップ 11 名の学生に関する興味分野のプロファイルを示す。これらの学生は貸出多数を占める学生の代表的存在であり、図書館にとって特に重要性の高い利用者である。順に P.A から P.K と名付け、氏名の直後のカッコ内の数字は貸出冊数である。所属に関しては理学部が 4 名、法学部が 3 名、文学部が 1 名、農学部が 1 名、その他が 2 名となっている。また、利用者タイプに関しては、学部学生が 4 名、修士学生が 2 名、博士学生が 4 名、その他 1 名である。

図より容易に分かることは、学生によって貸出カテゴリの割合が大きく異なることである。たとえば P.A の学生は極めて広範囲のカテゴリ (具体的には、000 から 900 の全てのカテゴリ) の図書を借りている。一方 P.C は対照的に極めて狭い範囲の図書を借りており、特に 400 (自然科学) 分野に集中している。

2.3 興味分野プロフィールの類似性指標

何らかの自動的手法により学習グループを発見するためのプロフィールの類似性を図る指標として本稿ではコサイン類似度を用いる。すなわち、2つのプロフィールベクトル $p = \langle p_0, p_1, \dots, p_9 \rangle$ と $q = \langle q_0, q_1, \dots, q_9 \rangle$ に対して次のように類似度を定める: $sim(p, q) = p \cdot q / |p||q|$, ここに $p \cdot q$

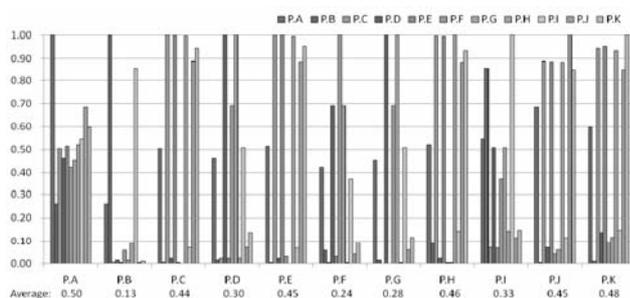


図 2: トップ 11 学生 P.A から P.K の相互類似度

はベクトル p と q の内積であり、 $|p|$ と $|q|$ はベクトルの長さ、すなわち 自分自身との内積の平方根である。コサイン類似度を用いる利点の 1 つは与えられたベクトルの類似度計算が極めて容易であることである。

図 2 にトップ 11 名の学生 P.A から P.K 相互の類似度を示す。自身との類似度常に 1 であり (i.e. $\text{sim}(p,p) = 1$), また、対称性がある (i.e. $\text{sim}(p,q) = \text{sim}(q,p)$)。図より P.A は比較的高い類似度をほとんどの学生に対して持つことが分かる。自身以外との類似度は P.J に対して最大であり (0.68), P.B に対して最小である (0.26)。類似度の平均値は 0.50 であり 11 名の学生中最大となっている。これは図 1 で示されるように P.A は広い範囲の興味を持つため、幅広い学生達と一定値以上の類似度を持つからである。

一方、P.B は他の学生達と類似度が P.I を除いて小さいことが分かる。P.I との類似度は 0.26 と比較的大きく、最小値は P.J に対する 0.01 であり、平均値は 0.13 である。本学生は他の学生とは異なり NDC カテゴリの 900 (文学) の図書を多く借りている。実際本学生は唯一の文学部所属の院生であり、恐らくは研究目的のために多くの文学作品を借りて読んでいるためと推察できる。

学習グループの候補を探すために他の学生との類似度が 0.9 以上の学生の組み合わせに着目する。その結果 4 名の学生が該当する。その結果を図 3 に示す。図から分かるようにこれら 4 名の学生間の類似度は全ての組み合わせで 0.9 以上になっている。すなわち、トップ 11 名の学生の中で、これら 4 名は興味分野の類似性に関して 1 つのクラスターを構成していることになる。更に詳しく調べると、

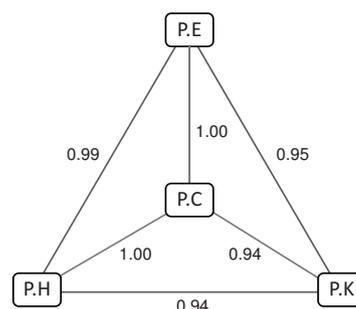


図 3: 類似度 0.9 以上の学生グループ

これらの学生達は、全て理学部/研究科に所属しており、学部 3 年生から修士の学生までの学生達である。修士の学生を中心に 1 つの学習グループとして共同で学習するのに、いかにも相応しいと考えられる組み合わせになっている。

3 利用者の専門度の指標

本稿では、論文 [5] で提案した専門度に関する 2 種類の指標のうち、c-rank と名付けた指標を用いて利用者の専門度を考察することにする。本指標は、専門度の高い利用者は専門度の高い図書を多く借りるものであるという仮定に基づいて定義される。一方、図書の専門度は、その図書を借りた利用者のタイプから想定される標準的な専門度に基づいて定義される。具体的には、利用者の想定専門度を次のように定める。まず、学部学生に関しては、その学年を専門度と定める。すなわち、学部 1 年次 (B1) から学部 6 年次 (B6) を、それぞれ専門度 1 から 6 を割り当てる。学習を主とする学部学生と、研究を主とする修士学生との差を強調するために、修士学生の専門度を 8 と定める。それに続いて博士学生には 9、教員には 10 を割り当てる。その他のタイプの利用者への専門度割り当てはなしとする。このような専門度の割り当てに基づき、図書の専門度を、その図書を借りた利用者タイプの専門度の平均値とする。このように定められた図書の専門度を基に、利用者の専門度 (c-rank) を、その利用者が貸出した図書の専門度の平均値として定める。

本定義は以下のように形式化できる。まず、図書の貸出データの集まりを CR (circu-

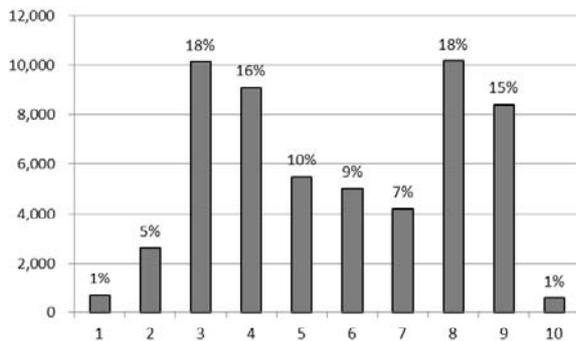


図 4: 全てのレコードの c-rank のヒストグラム

lation/loan records) と呼ぶ。それぞれの貸出データ (レコード) は次のような構造である。 $c = [bID = \text{図書の書誌 ID}, pID = \text{貸出者 ID}, pType = \text{貸出者の利用者タイプ}(B1, B2, \dots, B6, M, D, P, O), borrowDate = \text{貸出日}, borrowTime = \text{貸出時刻}, returnDate = \text{返却日}, returnTime = \text{返却時刻}, \dots]$ 本稿ではレコード c の bID (図書 ID) を $bID(c)$ と、そして pID (利用者 ID) を $pID(c)$ のように表現することにする。

利用者タイプの標準的専門度を $el(B1)$ や $el(P)$ のように表すことにすると、c-rank は以下の式で定義される。

$$c\text{-rank}(b) = \text{average}[el(pType(c))]: \\ c \in CR, bID(c) = b]$$

ここに *average* は平均値を表す。

図 4 に図書の専門度分布を示す。図によるとピーク値が 2ヶ所存在する。すなわち、3 と 8 である。学部 3 年次の学生が全体の 30% の図書を借りていることを考慮すると、ピーク値の 3 は 3 年次生を中心とした学部学生が主に借りている図書が該当するものと考えられる。一方、ピーク値 8 は、これらの図書が、特に修士学生を中心とした研究者レベルの利用者 (修士・博士学生および教員) に借りられていることを示していると考えられる。

4 学習グループ候補の選出

前節までの準備を踏まえて、本節では学習グループの候補を発見する手法を考察する。最初に、あ

るグループが学習グループとして適切であるか否かの判断基準を設定する必要がある。本稿では利用者と利用者の興味分野の類似度を第 1 の (最も重要な) 指標として採用する。それは、もしある学習グループのメンバが異なる興味範囲を持つならば、そのグループは同一の目標をもって協力しながらある教材を学習するのに適しないと考えられるからである。類似度の指標としてコサイン類似度 (Cosine Similarity) を用いる。

本節は以下のように構成される。まず第 4.1 節において、全ての利用者対に対してその興味分野の類似度を計算する。次に類似度に基づいて、類似度の高い利用者からなる学習グループ候補の可能性を検討する。その結果、類似度の高さだけでは学習グループとして不適切と考えられる所属する人数が多すぎるグループが存在することが判明した。そのようなグループに関しては、更に進んだ検討が必要となる。

次に第 4.2 節では、NDC カテゴリのトップレベルではなく、そのサブカテゴリを用いることにより、詳細化された興味分野を用いることで、類似性の詳細化を試みる。その結果、あるグループは適切なサイズに落ち着くものの、そうでないグループも存在することが分かる。それらのグループに対しては更なる分割の方策が必要となる。

第 4.3 節では、専門度を併せて用いることにより、グループを細分化することを試みる。このような方式を採用することにより、より妥当な方法で適切なグループを構成することができる。

最後に、第 4.4 節で、更に優れた学習グループ構成手法に関して検討し、本節の議論をまとめる。

4.1 興味分野の類似度を用いた分析

本節では、5,718 名の利用者の興味分野の類似度分析を行う。すなわち、16,344,903 対の類似度をまず計算する。図 5 に、類似度のヒストグラムを示す。x 軸の 0 と 1 は、それぞれ類似度が正確にその値である場合である。それ以外は、小数点以下 1 位を四捨五入した値である。

図から分かるように、ほとんどの対 (9,072,290, 55.5%) は類似度 0 である。逆に、1,383,927 対

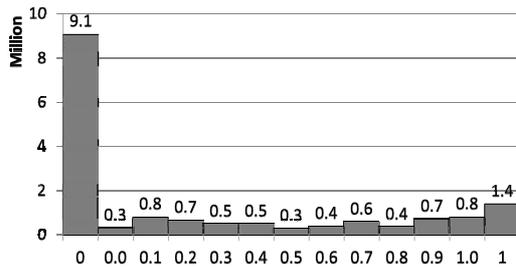


図 5: 全ての利用者の類似度のヒストグラム

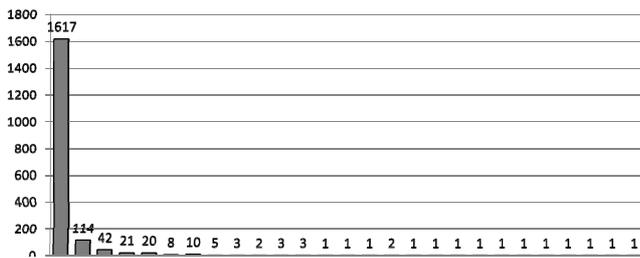


図 6: グループサイズに対する該当グループの頻度

(8.5%)は類似度1である。利用者 A と利用者 B の類似度が1であることを $A \sim B$ と表すと、この関係は同値関係となる。すなわち、(i) $A \sim A$ (反射性), (ii) $A \sim B$ ならば $B \sim A$ (対称性), そして (iii) $A \sim B$ かつ $B \sim C$ ならば $A \sim C$ (推移性) の3つの性質を満足する。したがって、本関係 \sim を用いた同値クラスによって全ての利用者をクラス分けすることができる。

図6はx軸にグループサイズをy軸にそのサイズのグループ数を示す。グループのサイズは1から最大1,383という巨大グループと広い範囲にわたる。図より、サイズが1のグループが1,617もあることが分かる。そのようなグループのメンバは、全く同じ興味分野のプロファイルをもった利用者が他にいない独自のプロファイルを有していることになる。同様にメンバ数が2のペアを成しているグループは114存在し、また、3名のメンバからなるグループは42あることが分かる。このようなメンバ数を持つグループは、学習グループとしてはサイズが小さすぎる恐れがあり、興味分野の類似性の範囲を広げることにより、より大きなサイズにすることで、適切なサイズの学習グループ候補の生成を考える必要がある。

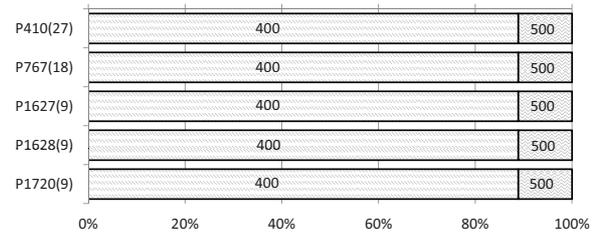


図 7: あるサイズ5グループのメンバのプロファイル例

学習グループのメンバ数は重要である。もし学習グループの人数が多すぎた場合、メンバの興味や共同学習のやり方に関する意識のばらつきが大きくなるなどの問題が発生し、その結果全てのメンバが十分に学習できるようにするのが困難となる。一方人数が少なすぎた場合、たとえば、ある特定のメンバが一方向的に教える立場になったり、逆にあるメンバが教えられるばかりになったりして他のメンバの足を引っ張るような状況も考えられる。理想的な学習グループとしては、全てのメンバが、別の内容に関しては教える役割となり、別の内容に関しては教わる役割となり、両者のバランスが上手く取れているものが相応しいと考えられる。少人数のグループの場合、このバランスが崩れる恐れが高い。したがって、上手くマッチングするという見込みがある場合を除いて、少なすぎるメンバによる学習グループの構成は避けるのが賢明であると考えられる。最適なグループサイズは、状況に応じてさまざまであり、一概には言えないものの、我々の経験を踏まえると、一般的な状況としては、本稿では4名から6名程度を最適なサイズとするのが妥当ではないかと考えられる。以下、最適サイズを4~6名、最大数8名として考察を進める。

図6によると、64グループ、人数では342名がサイズ4から8までに含まれ、本稿の条件にマッチする。図7にサイズが5である、あるグループのプロファイル例を示す。図より、410番目から1720番目の全てのメンバが類似した興味分野のパターンを持っていることが分かる。これらの学生達の貸出数は27冊から9冊とばらつきはあるものの、これらの学生達は全てNDCカテゴリ400(自然科学)と500(技術)のみからの図書を8対1の

比で借りている点で共通している。グループサイズ1から3 (1,971名の利用者) に関しては、類似度が高い学生から1名から5名のメンバを追加することにより、条件にかなったグループを構成することができる。

8名を超えるサイズを持ったグループ (9名から1,383名) に関しては、最適と思われるグループを構成することは、もっと困難である。これらのグループのメンバは貸出冊数は異なっているとしても、まったく同じパターンで図書を借りている。これらのグループに属する学生達に適したグループ構成を推薦するためには、グループをランダムに分割するか、別の基準を導入することにより、より適切なグループの構成を目指すか、2つの選択肢がある。新しい基準の1つの案は、NDCトップレベルのカテゴリを詳細化した類似度尺度を導入し、更に細かい興味分野の類似度を調べることである。このアプローチは次節、第4.2節で追究する。もう1つの案は、興味分野とは異なるタイプの指標を導入することである。その候補として、利用者の専門度を用いたアプローチについては、第4.3節で議論する。

4.2 興味分野の更なる詳細化によるグループ候補探索

サイズの大きなグループを分割するための妥当な手法を開発するために、最大のグループを子細に分析することにする。本グループは貸出冊数順で14番目 (127冊) の学生から5,715番名の学生 (1冊) にわたる1,383名のメンバから構成されている。貸出冊数は異なるものの、これらの学生は全て、NDCカテゴリ400 (自然科学) 分野の図書のみを借りている点で共通している。

2番目に大きなグループは76番目 (83冊) から5,716番目 (1冊) にわたる846名からなり、NDCカテゴリ300 (社会科学) の図書のみを借りている学生により構成されている。他のグループを見ると、その他の大人数のグループも同様に限られた興味分野、すなわち特定のNDCカテゴリ、の図書のみを借りている学生達から構成されている。たとえば、3番目に大きなグループはNDCカテゴ

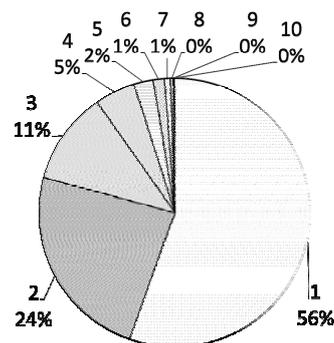


図8: 貸出図書のNDCカテゴリ数に対する利用者の割合

リ500 (技術) のみであり、4番目は900 (文学)、5番目は800 (言語) といった具合である。

図8に利用者の貸出NDCカテゴリ数 (1から10) に対する利用者数の関係を示す。図よりおよそ60%の利用者がただ1つのNDCカテゴリの図書のみを借りていることが分かる。ただ1冊の図書のみを貸出した学生は当然ながらNDCカテゴリ数も1となるため、60%の中にはそのような学生も含まれることに注意が必要である。同様に約1/4の学生は2つのNDCカテゴリの図書のみを借りている。また、1/9の学生は3つのカテゴリの図書のみを借りている。図1よりP.AからP.Kのトップ11名の学生達のプロフィールを改めて見ると、P.CとP.Hはわずか3つのNDCカテゴリのみから図書を借りている。他の9名の学生は、10種全てのNDCカテゴリから借りているP.AとP.Kを含め、4つ以上のカテゴリの図書を借りている。多くの図書を借りている学生達は全体として多くのカテゴリの図書を借りており、一般の学生達と比較して興味分野が広いことが窺える。

ここでメンバ数最大のグループを分析する。本グループはNDCカテゴリ400 (自然科学) の図書のみを借りている学生により構成されているグループである。Grp400と名付けよう。Grp400に対する最初の分析として、これまでのNDCカテゴリを用いた分析法と同様に、NDCカテゴリを更に細分化した興味分野プロフィールにより類似度を調べる。すなわち、NDCカテゴリ400を十位で展開したNDCカテゴリ400から490までの図書の貸出冊数を基に細分化されたプロフィールに対

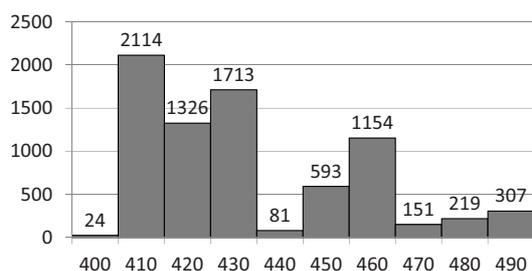


図 9: Grp400 のメンバにおける 400 から 490 までの NDC サブカテゴリに関する貸出図書の延べ数

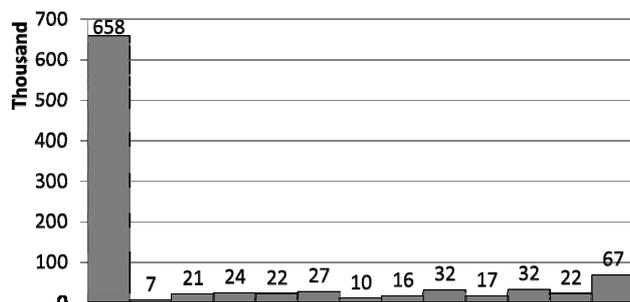


図 10: NDC400 カテゴリの図書のみを借りた学生達の類似度に関するヒストグラム

する同様の類似度により第 2 段階のグルーピングを行う。図 9 に Grp400 のメンバに関する NDC カテゴリ毎の貸出冊数のヒストグラムを示す。本グループを構成する学生が借りた図書は総数で 7,682 冊である。図より、最も多く貸し出されたカテゴリは 410 (数学) であり、それに引き続き 430 (化学), 420 (物理学), 460 (生物学), 450 (地球科学) となっている。

NDC のトップカテゴリの場合と同様な方法により、Grp400 に属する 1,383 名のメンバに対する 955,653 対の類似性を計算した結果のヒストグラムを図 10 に示す。図より、トップレベルのカテゴリに対すると同様に、69% (657,799 対) が類似度 0 であり、7% (67,293 対) が類似度 1 である。トップレベルの場合と同様に、類似度 1 なる関係を用いて Grp400 のメンバをクラス分けする。その結果、最大のグループとして NDC カテゴリ 420 (数学) の図書のみを借りた学生からなるメンバ数 205 のグループが得られる。

類似度を比較するためのカテゴリを細分化する

ことにより、元のグループと比較すると確かに小さなグループを得ることができた。しかし、205 名からなるグループは学習グループとしては依然大き過ぎるグループとなっている。それに対する対策としては、NDC カテゴリを更に 1 の位まで細分化することが考えられる。さらには、小数点以下にまで細分化を進めることも手法としては可能である。しかし、一方では、あまりに下位レベルにまで細分化してしまうと、利用者の貸出履歴のわずかの差異で、同じ学習グループのメンバとして興味分野の差異があるものと判断されることが生じ、判断基準として適切ではなくなる恐れがある。

したがって、NDC カテゴリを用いた興味分野の把握としては、1 位の NDC カテゴリ程度 (3 段階) に留めておくのが妥当であるものと考えられる。したがって、それ以上のグループの細分化手法としては、NDC カテゴリ以外の判断基準を導入するのが望ましい。その 1 つとして次節 4.3 では専門度によるグループ分けの手法について議論する。

4.3 専門度も加味したグループ分け手法

本稿では、NDC 番号を用いた興味分野のパターンの類似度指標に基づいて利用者のグループ分けを行ってきた。しかし、生成されたグループ中の一定割合はグループサイズが大きく、学習グループの候補を構成するには精度が十分ではない。本節では、第 3 節で説明された専門度を新たな指標として導入することにより、興味分野に基づいて得られたグループ候補を細分化することを試みる。

NDC カテゴリのトップレベルを用いたグループ分けにより得られた最大のグループは、NDC400 (自然科学) の図書のみを借りたグループ Grp400 であり、それを NDC の十位のカテゴリにより細分化して得られたグループの中のサイズ最大なのは NDC カテゴリ 410 (数学) の図書のみを借りるグループであった。

このグループに属する利用者の中で貸出数順序が最小 (すなわち貸出数最大) の学生が 52 番目の学生 (貸出数 78 冊) であるため、この学生を代表として本グループを G52 と呼ぶことにする。G52

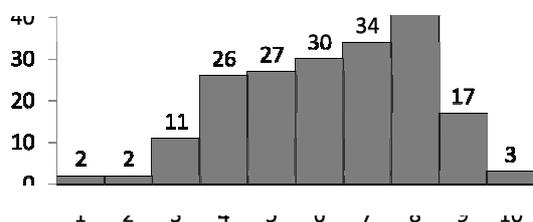


図 11: G52 グループの c-rank 値に関する利用者数

グループは 205 名の学生から構成される。

図 11 に G52 グループに所属する学生達の専門度 c-rank の値 (1 から 10) のヒストグラムを示す。NDC カテゴリと異なり、専門度が完全に一致するケースはまれであるため、専門度の近い利用者からなる学習グループを適切な人数 (サイズ) により構成することは容易である。

G52 に所属する学生を専門度により細分化すると以下ようになる: 最初のグループは専門度が 9 以上の 7 名のメンバから構成される。次のグループは残りの学生の中から専門度が 8.6 以上 9 未満の 6 名, それに引き続き専門度 8.2 以上 8.6 未満の 6 名となり, 以下同様の手続きで, 目標サイズのグループを構成していくことができる。

4.4 更なる分析手法に関する議論

本節ではまず「どういう学生グループが学習グループとして一緒に学習するのに適しているか?」そして「いかにして貸出データから適切なグループを見つけることができるか?」という基本的問いかけへの解答を議論する。これらの質問は非常に基本的であると同時に、解答を見つけることは極めて困難である。本稿では本質問に対する最終的な解答を追究する前に、まず第 1 段階として、必要条件を考察する。

少なくとも、グループメンバの興味分野が大きく異なれば、一緒に学ぶ教材の選定において問題が生じることが予想できる。そのため、本稿では第 1 の条件として同じ学習グループのメンバは共通の興味分野を有することを課した。それと同時にメンバの知識レベルに関しても同程度レベルで

あることを条件とする。そうすることにより、一部のメンバは非常に高い知識レベルを持ち、したがって他のメンバにもっぱら教える役割を果たすことになったり、逆に一部のメンバの知識レベルが低すぎて他のメンバの足を引っ張るような状況になったりする状況をできるだけ回避したい。我々はすでに論文 [5] において図書館利用者の専門度 (知識レベル) を推定する方法を提案している。本稿では、その結果を利用する。

グループサイズ (所属メンバ数) も議論すべき重要なテーマである。もしグループのサイズが大きすぎるならば、そのグループは共同で学習するためのグループとして効果的でないものと考えられる。我々の経験を踏まえると 4 から 6 名からなるグループが通常の状態において適切なグループサイズであると考えられる。もっとも適切なグループサイズはどのような内容を学ぶのか、また、どのようなやり方で学ぶのか等に応じて多少変化するものと考えられる。更に詳しい検討が必要である。本稿では、最適なグループサイズを 4 から 6 名、多くても 8 名程度と設定した。

我々の当初の考えでは、クラスタリング技術が学習グループ候補の生成に適用可能であると判断していた。しかし、それは常に適切な手法とは限らない。もし学生がいくつかの主題 (トピック) に関する学習グループに所属したいと希望した場合、それらの学習グループは同一時間帯に設定することはできない。

したがって、通常の (ハード) クラスタリングのように 1 人のメンバが 1 つのグループに所属するような配分は必ずしも適切ではないことになる。一方、少人数のグループに分かれて同じ時間帯に学習会を開催するような状況においては、このクラスタリング手法はグループ候補生成に有効な手法ということになる。

もっと一般的な状況においては、1 人の学生は 2 つ以上のグループに所属することができ、所属しているそれぞれ学習グループにおいて、それぞれ異なるトピックを学習することが可能であるような学習グループの割り当てが必要である。このような一般的な状況における学習グループ候補の生成に関しては将来の課題とし、本稿では扱わない。

5 まとめの議論と将来への課題

本稿で提示された研究の主な動機や目的としては(1)新しい図書館サービスを追究し、激しく変化しつつある社会の中で、図書館が今後も有益な社会的役割を果たし続けられるようにすること、(2)図書館マーケティングのための基礎となるデータとして、貸出データが有効であることを示すことを、特に学習グループの候補生成を通じて示すことがある。また、学習グループ候補を推薦するサービスは現在のところ図書館サービスとして一般に認知されていないものの、LCの有効利用に資することなど、今後の図書館における新サービスとして高い潜在力を持つことを示すことも重要な目的と言える。

本稿では、我々は大学図書館の貸出データを用いることにより、大学における学生への学習グループ推薦への適用を念頭に議論を展開した。そのために貸出データから得られる利用者の興味分野のプロファイルや、同じく貸出データから得られる利用者の専門度指標を用いるグループ構成法を示した。その結果、本稿で提案する方法により学生に対する学習グループ候補の推薦が効果的に行えることを示した。

本稿で示したような手法のみでは、得られた学習グループ候補が所属している学生達にとって本当に適しているグループになっているか否かの最終的判断は困難である。実用的な学習グループを構成し、実施するためには、適性のマッチングに関して更なる検討が必要である。とはいうものの、白紙の状態から、本稿で示したような方法によって、(半)自動的に学習グループの候補を選定することにより、最終的には人手による検討を行うとしても、その手間を大幅に減少させることが可能である。また、興味分野や知識レベルに関して大きなアンバランスのあるグループを構成してしまうリスクを少なくすることができる。

今後、本稿で示された手法を更に改善するとともに、次のようなトピックに挑戦することが重要である:(1)実用的な学習グループの構成に役立つであろう更なるパラメタの発見とその適用方法に関する研究、(2)興味分野に関する追加情報の追究および、更に効果のありそうなパラメタの発見

とその組み合わせ方に関する研究、(3)本稿で示したモデルを拡張することにより、より多様な学習内容や学習スタイルに適合した学習グループの発見手法に関する研究、(4)本稿の研究および今後の研究の結果得られたグループ候補に対して、実際に学習グループを構成し、実践し、評価する実験を行い、真に有効な学習グループの構成に必要な新たなパラメタを発見し、また、どのような工夫を行うことでメンバにとってより有効な学習成果が得られるようにできるかに関する研究、等。

図書館マーケティングの一環として、学習グループの候補を発見し、推薦するという、本稿で示した研究テーマは、従来なされてきた学習環境の整備による学習支援から、図書館によるより能動的な学習支援を目指した新しい研究分野/テーマである。本稿で示した研究成果は、その目標に対して、まだまだ十分なものではない。しかし、その重要性を考えると、今後多くの研究がなされるべきであり、将来的には図書館の利用者サービスとして必要不可欠なものとなることが期待できる。

謝辞

本研究の一部は文科省の科研費、基盤研究(C)、24500318, 2013, および九州大学P&Pとして実施された。

参考文献

- [1] Chen, C., Zhang, L., Qiao, H., Wang, S., Liu, Y., Qiu, X.: Book Recommendation Based on Book-Loan Logs. In: Chen, H.-H., Chowdhury, G. (eds.) ICADL 2012. LNCS, vol. 7634, pp. 269–278. Springer, Heidelberg (2012)
- [2] Konstas, I., Stathopoulos, V., Jose, J.M.: On Social Networks and Collaborative Recommendation. In: 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 195–202. ACM Press, New York (2009)

- [3] Li, H., Gu, Y., Koul, S.: A Review of Digital Library Book Recommendation Models. Available at SSRN: <http://ssrn.com/abstract=1513415> (2009)
- [4] Littman, J., Connaway, L.S.: A Circulation Analysis of Print Books and e-Books in an Academic Research Library. *Library Resources & Technical Services*, 48(4), pp. 256–262 (2004)
- [5] Minami, T.: Expertise Level Estimation of Library Books by Patron-Book Heterogeneous Information Network Analysis – Concept and Applications to Library’s Learning Assistant Service –. The 8th International Symposium on Frontiers of Information Systems and Network Applications (FINA 2012), DOI 19.1109/WAINA.2012.184, pp. 357–362 (2012)
- [6] Minami, T., Baba, K.: Investigation of Interest Range and Earnestness of Library Patrons from Circulation Records, International Conference on e-Services and Knowledge Management (ESKM 2012), as a part of the 1st IIAI International Conference on Advanced Applied Informatics (IIAI-AAI 2012), IEEE CPS, DOI 10.1109/IIAI-AAI2012.15, pp. 25–29 (2012)
- [7] Minami, T.: Profiling of Patrons’ Interest Areas from Library’s Circulation Records – An Approach to Knowledge Management for University Students –, The Fifth International Conference on Information, Process, and Knowledge Management (eKNOW 2013) (2013)
- [8] Minami, T.: Interest Area Analysis of Person and Group Using Library’s Circulation Records, Proc. IADIS International Conference Information Systems (IS 2013), 8pp. (2013)
- [9] Minami, T.: Changes of Interest Range of Students with Circulation Record Analysis, Proc. Information Conference 2013 (Information’13), 4pp. (2013)
- [10] Minami, T. and Baba, K.: An Attempt to Find Potential Group of Patrons from Library’s Loan Records, International Symposium on Advanced and Applied Convergence (ISAAC 2013), 4pp. (2013)
- [11] Minami, T. and Baba, K.: LIBRARY PATRON’S INTEREST AREA PROFILING AND SIMILARITY BASED GROUPING – WITH AN INVESTIGATION ON APPLICABILITY TO STUDY GROUP FORMATION –, International Conference on Applied and Theoretical Information Systems Research (ATISR 2013), 11pp. (2013)
- [12] O’Hearn, K.: Computer-Aided Student Study Group Formation, *McNair Scholars Journal*, vol. 13, iss. 1, Article 10 (2009) <http://scholarworks.gvsu.edu/mcnair/vol13/iss1/10>
- [13] Ranganathan, S.R.: The Five Laws of Library Science, Bombay Asia Publishing House (1963)
- [14] Resnick, P., Varian, H.R. (Guest Eds.): Recommender Systems, *Communications of the ACM*, vol. 40, no. 3, pp. 56–58 (1997)
- [15] Tsuji, K., Kuroo, E., Sato, S., Ikeuchi, U., Ikeuchi, A., Yoshikane, F., Itsumura, H.: Use of Library Loan Records for Book Recommendation. In: 3rd IIAI International Conference on e-Services and Knowledge Management (IIAI ESKM 2012). DOI:10.1109/IIAI-AAI.2012.16 (2012)