

ソーシャル・デジタルライブラリにおける

日本十進分類法を用いた人物推薦手法

筑波大学

図書館情報メディア研究科

2014年5月

常川真央

概要

題目：ソーシャル・デジタルライブラリにおける日本十進分類法を用いた人物推薦手法

本研究の目的は、ソーシャル・デジタルライブラリに適した人物推薦手法を提案することである。デジタルライブラリはユーザが属するコミュニティの社会的文脈を反映し、新たな社会的相互作用を生み出す媒介物として機能すると様々な図書館情報学者から指摘され、協調的な知識コミュニティとしての側面の研究や概念モデルの提示も行われてきている。Worrall は、ユーザ同士のコミュニケーションの媒介物となり、コミュニティ形成の機会を提供する機能を持つデジタルライブラリをソーシャル・デジタルライブラリと呼び、調査に基づいて詳細な役割を明らかにした。また、コミュニケーションの仲立ちとなる役割を、Worrall はバウンダリオブジェクトという概念を援用して明確化した。バウンダリオブジェクトとは、ユーザ同士のコミュニケーションの媒介物となり、新たなコミュニティを形成する役割を担う組織・用語・事物を指す概念である。Worrall の研究によって、ソーシャル・デジタルライブラリの重要性は明らかになった。しかし、バウンダリオブジェクトが具体的に何を指すのか、そして、コミュニティ形成を支援する機能の実装手法については、未だ具体的に検討されていない。

そこで本研究では、ソーシャル・デジタルライブラリにおけるコミュニティ形成を支援するための人物推薦手法を提案する。デジタルライブラリに登録されているユーザの資料利用履歴から、分類番号を抽出・類型化することでユーザプロフィールを作成し、類似

した関心を持つ他のユーザを発見するアルゴリズムを考案することで、研究目的を達成する。

本論文は6つの章によって構成されている。

第1章では、本研究の背景と目的を記述する。

第2章では、関連分野の研究動向をソーシャル・デジタルライブラリとオンラインコミュニティ、人物推薦システムの3つの研究領域に分けて記述する。既存のソーシャル・デジタルライブラリは、ソーシャル OPAC とソーシャルカタログ・サイト、読書系 SNS の3つの関連領域の中で別々の取り組みとして企画・運用されている。関連領域それぞれのソーシャル・デジタルライブラリは、機能には差異はみられるが、特定の情報資源を媒介としてユーザ同士のコミュニケーションを行うことを支援するという点で共通している。このように、ソーシャル・デジタルライブラリは限定的ではあるがシステムとして開発・運用されている。コミュニケーション支援の課題の1つとして、ユーザ同士の出会いを直接支援する類似読者推薦機能の実現がある。類似読者推薦機能とは、ユーザが登録した資料利用履歴に基づき、ユーザにより近い関心を持つ他のユーザを検索し、コミュニケーションの相手として推薦する機能である。類似読者推薦機能は LibraryThing や読書メーターなど様々なソーシャル・デジタルライブラリに実装されており、ユーザ間のコミュニケーションの直接の機会を提供している。ただし、類似読者推薦機能が基準とする読書傾向の類似性の尺度の妥当性については、明確には研究が為されていない。

オンラインコミュニティの研究は、インターネットが発展した初期の段階から様々な研究領域を横断して研究されてきた。その結果、共通の関心に基づくオンラインコミュニティをいかに形成していくかが重要であることが明らかになっている。

人物推薦システムは、ユーザがコミュニケーションを取りたいと思う人物を推薦するシステムであり、ユーザの嗜好を推測して能動的にコンテンツを推薦する推薦システムの一

領域として研究が開始された。その後、人間関係を支援する情報システムの実現という独立した研究領域として位置づけられ、研究開発が盛んになされるようになった。人物推薦システムのアプローチには、共通の交友関係に基づく Friend-of-Friend 方式と、ユーザの投稿内容の類似性に基づく Content-Matching 方式の 2 種類がある。前者は知人との交流を促進するために、後者は共通の関心を持つ人物と出会うことを目的とする状況で有効であることが明らかになっている。

以上の関連研究から、本研究の課題は共通の関心に基づくコミュニティを形成するための人物推薦手法を提案することであり、そのためには Content-Matching アプローチの人物推薦手法を採用することが望ましいことを明らかにしている。また、Content-Matching アプローチの実現には、ユーザプロファイリング技術と人物推薦アルゴリズムの 2 つの技術を提案する必要があることも明らかにしている。

第 3 章では、ソーシャル・デジタルライブラリにおけるユーザプロファイリング技術として、NDC ツリープロファイリングを提案する。ソーシャル・デジタルライブラリにおいて、ユーザの関心を推測する手がかりとなるものに資料利用履歴がある。資料利用履歴とは、ユーザの貸出履歴や読書履歴、検索履歴などのように、ソーシャル・デジタルライブラリが保有する資料に紐付けられたユーザの行動履歴であり、資料の目録情報と資料に対するユーザの行動情報によって構成されている。ここで、資料利用履歴に含まれる資料から、資料の主題を表わす分類番号を抽出・集計することで、利用した資料に関連したユーザの関心を明示化することができる。特に国内で多くの図書館に利用されている分類法である日本十進分類法 (NDC) は、階層構造の分類体系であり、本手法に適している。そこで、本研究ではユーザの資料利用履歴から NDC 番号を抽出し、NDC ツリープロフィールと呼ぶ、ユーザの読書傾向を表現するツリー状のユーザプロフィールを作成する手法として、NDC ツリープロファイリングを提案する。

第4章では、ソーシャル・デジタルライブラリの人物推薦アルゴリズムを提案する。それに伴い、ユーザプロフィール同士の類似性を測るために、形状類似スコアと共通ラベルスコアの2種類の尺度を提案する。さらに、2つの尺度を補完する尺度としてコメント率を提案する。そのうえで、人物推薦を実現するための推薦スコア算出アルゴリズム NDCTREE1 を提案する。さらに、提案アルゴリズムの有効性を検証するために、利用者実験を行い、その結果から得た知見をもとに改良したアルゴリズム NDCTREE2 を提案する。NDCTREE2 には、4種類のパラメータを設定している。まず、共通ラベルスコアに対するパラメータとして w_1 と w_2 , w_3 を設定しており、プロフィールの類似性を算出する際に重視する NDC ツリープロフィールの階層を決定する。次に、形状類似スコアと共通ラベルスコアのバランスを調整するためのパラメータとして $u(0.0 \leq u \leq 1.0)$ を設定している。NDCTREE2 について、再度評価実験を行うことでその有効性を検証した。その結果、推薦の再現率は 0.3382、精度は 0.3473 であり、3つの比較手法である RANDOM・VECTOR・BOOK より優れていることが明らかになった。ただし、実験協力者の資料利用履歴の大半は小説が占めていた。この影響を確認するために、小説を抜いたデータで評価実験を行ったところ、再現率 0.3618、精度 0.3052 であることが分かった。

第5章では、第4章で記述した評価結果から、提案手法の考察を記述する。本研究は、人物推薦の新たな手法として NDC ツリープロファイリングを用いた人物推薦手法を提案し、次のような有効性が示された。まず、本手法のほうが RANDOM よりも再現率・精度が高いことが示された。このことから、本手法は人物推薦としての意義があることが分かった。さらに、主題の階層性を考慮しない推薦方式である VECTOR と BOOK よりも再現率・精度が高いことが示され、パラメータの最適化において NDC の第二次区分に対応する w_2 を重視することで結果が最適化されることが示された。このことから、階層的な主題の共通性が人物推薦に良い結果をもたらすという本研究の仮説が検証された。

さらに、小説を抜いたデータによる評価実験の結果によって、本手法の可能性を明らかにした。小説を抜いたデータによる評価結果において、小説を含めた場合と同等の性能に最適化することができた。これは、重要視する主題の粒度を調節し、最適な人物推薦を実現できる可能性があることが示された。このことから、本手法では利用する資料利用履歴の傾向に応じて主題の粒度を調節することができることが分かった。

一方で、本手法の問題も明らかにした。形状類似スコアの重みは 0.1 とわずかであることから、全体的な傾向として形状類似スコアが人物推薦の性能に与える影響はわずかであった。ただし、個別の実験協力者を見たところ、形状類似スコアを重視することにより F 値が上がる実験協力者もいることが明らかになった。また、形状類似スコアが有効な実験協力者からは関心の幅が狭い NDC ツリーを持つという傾向が観察され、形状類似スコアが有効でない実験協力者からは関心の幅が広い NDC ツリーを持つという傾向が観察された。形状類似スコアと NDC ツリーの関係性について検証する必要がある。

第 6 章では、結論を述べる章である。第 5 章の考察から、NDC ツリープロファイリングを用いた人物推薦手法は人物推薦として意義のある手法であることが明らかになった。さらに、本研究の評価の範囲では、ユーザ間の関心の類似度を測るには、階層的な主題が適切であることが分かった。このことは、ソーシャル・デジタルライブラリにおけるバウンダリオブジェクトが具体的に何かを示唆する知見として役立つ。また、人物推薦アルゴリズムで用いる尺度については、形状類似スコアの有効性はユーザの関心の広がりに関連があることが分かった。本論文で明らかになった結果は、限定的な実験環境下から導かれたものである。今後の研究課題として、実験条件の改善と、形状類似スコアの影響の詳細な分析を通じた算出式の更なる改善が挙げられる。以上の課題を解決することで、ソーシャル・デジタルライブラリにおけるコミュニティ形成の包括的な支援へ研究が発展していくことが考えられる。

Abstract

Title: A Method of Person Recommendation with Nippon Decimal Classification in Social Digital Library

This study proposes a person recommendation method for supporting community creation in social digital libraries. Various scholars of library and information science have pointed out that digital libraries encourage social interaction among digital library users, reflecting the social contexts of the community to which digital library users belong. They also propose various concept models of digital libraries as collaborative communities of knowledge. Worrall described digital libraries offering community-creation opportunities as “social digital libraries,” and he defined their detailed roles based on surveys. He also clarified the roles of communication bridges played by digital libraries aided by the concept of “boundary objects.” Boundary objects include things, organizations, and terms, that facilitate communication among users and act to create new communities. Worrall’s research emphasized the importance of social digital libraries. However, the meaning of boundary objects in social digital libraries and ways to implement functions that support community creation have not been studied yet.

In this paper, we propose a method for person recommendations for supporting

community creation in social digital libraries. Therefore, we propose a user profiling method with classification numbers extracted from resource use logs and a recommendation algorithm for finding users with similar interests.

This paper consists of six chapters. In Chapter 1, we describe background and purpose in this study. In Chapter 2, we separately discuss three research trends, namely, social digital libraries, online communities, and a person recommendation system. Social digital libraries are being planned and used in different related fields, namely, social online public access catalogs (SOPACs), social cataloging, and social networking sites (SNSs) for reading. In each field, social digital libraries serve a different function, but they are common in that they support communication among users through an information resource. As one of the subjects of communication support, there is the realization of a “similar reader recommendation function” that facilitates user interaction. The similar reader recommendation function searches for users with similar interests based on their resource use logs and recommends them to the target users as potential partners for communication.

Some existing social digital libraries such as LibraryThing or Dokusho-meter have similar reader recommendation functions for offering users the opportunity to communicate with other users. However, little is known about what is the appropriate measure for calculating similarities between the reading tendencies of users.

Online communities have been studied across various research areas since the beginning of the Internet. As a result, it is clear that the creation of communities based on common interests is an important step in the growth of online communities.

Investigation of person recommendation systems, which recommend persons who

are expected to benefit from communication with each other, was started as a sub-field of recommendation systems. Later, it was addressed as an independent field by Terveen, and has been growing and has been vigorously investigated. There are two principal approaches for a person recommendation system. The first approach is called the “friend-of-a-friend” approach, which is based on a friendship network among users. This approach is effective for supporting interaction among users and their friends. The second approach is called the “content-matching” approach, which is based on similarity of content posted or accessed by users. This approach is effective for supporting interactions among users who have similar interests.

As mentioned above, our research is on ways to implement a person recommendation system for creating online communities of people with common interests in social digital libraries; therefore, it is appropriate for this system to adopt the content-matching approach. Consequently, the concrete subject of our research is to develop a user profiling technique and a person recommendation algorithm for realizing the content-matching approach.

In Chapter 3, we propose using Nippon Decimal Classification (NDC) Tree Profiling as the method of user profiling in social digital libraries. This paper focuses on resource use logs as key data for determining the interests of users. A resource use log is the history of a user’s library resource access in the social digital library, such as the history of library loans, books read, or searches. The log consists of catalog records of the target resource and information about the user’s behavior toward them. We can clearly express a users’ interests by extracting classification numbers from their resource use logs.

Particularly, NDC is suitable for our method because it is used by a large number of libraries in Japan. Thus, we propose a user profiling method using NDC called “NDC Tree Profiling.” In this method, the recommendation system creates a user profile called an “NDC Tree Profile”, which uses a tree structure to represent reading tendencies of users.

In Chapter 4, we propose a person recommendation algorithm for social digital libraries. In this algorithm, we define two measures: “form similarity score,” “common label score”, in order to calculate similarities among user profiles. We also define a measure: “comment rate”, in order to complement other two measures. Then, we propose NDCTREE1, which is a score calculation algorithm for generating person recommendations based on the above three measures. We also propose NDCTREE2 as an improved algorithm through experimental results of evaluation of NDCTREE1. NDCTREE2 has four parameters: w_1 , w_2 , w_3 , and u . The parameters w_1 , w_2 , and w_3 are used for the common label score. When the recommendation system identifies similarities among user NDC tree profiles, these parameters determine which layer the recommendation system regards as important. The parameter u is used for balancing the form similarity score and the common label score. We carried out an evaluation experiment to assess the usefulness of NDCTREE2 by repeating the experiment. As a result, we obtained recall and precision scores of 0.3382 and 0.3473, respectively, which made it clear that our method is more efficient than RANDOM, VECTOR, and BOOK. However, the majority of resources in examinees’ resource use logs are novels. For verifying the influence of this deviation, we performed an additional experiment using a subset of the resource use logs that excluded novels. As a result, we obtained

recall and precision score of 0.3618 and 0.3052.

In Chapter 5, we present a discussion about the experimental results of our method. In the experimental results, the usefulness of our person recommendation method using NDC Tree Profiling was shown as follows. First, experimental results show that the recall and precision of our method are higher than RANDOM. This shows that our method might be useful for person recommendation. Second, the recall and the precision of our method are higher than VECTOR and BOOK, which are person recommendation methods that do not consider a hierarchy of subjects. Consequently, we have proven our hypothesis that the common ground of hierarchical subjects is effective for improving the precision of person recommendation.

Moreover, we made clear the potential of our method with experimental results using a subset of data that excluded novels. The best result using this subset data achieved performance equivalent to that using total data. It is realized by changing the granularity of subjects that recommender system regard as important. This showed that by adjusting the granularity of the subjects according to the reading tendencies of users, an optimal person recommendation could be realized by our method.

However, problems with our method were also clarified. Because the optimal weight of the form similarity score is 0.1, the form similarity score had little effect on the efficiency of person recommendation. The form similarity score is useful for examinees who have narrower NDC trees, but is not useful for examinees who have broader NDC trees. Therefore, our method needs to be further improved by adjusting the weight of the form similarity score for each examinee. Additional verification is also necessary under improved experimental conditions.

In Chapter 6, we present the conclusion. Discussions in Chapter 5 made it clear that our person recommendation method using NDC tree profiling is useful. Moreover, from the experimental results, it was found that hierarchical subject areas are appropriate for measuring similarities of interests among users. This result is useful for defining boundary objects concretely in social digital libraries. To measure the accuracy and effectiveness of the person recommendation algorithm, we found that the form similarity score closely relates to the range of a target user's interests. In conclusion, the experimental results made it clear that our method will significantly contribute in the development of social digital libraries. However, our conclusion was introduced from limited experimental conditions. Improving experimental condition and score calculation algorithm is necessary for the further research. If these problems have been solved, our study will develop into a study for comprehensive supporting community-creation in social digital libraries.

目次

第1章	はじめに	1
1.1	デジタルライブラリの社会的文脈	1
1.2	ソーシャル・デジタルライブラリ	2
1.3	本論文の構成	3
第2章	関連研究	7
2.1	ソーシャル・デジタルライブラリの現状と課題	7
2.2	オンラインコミュニティの様態	14
2.3	人物推薦システム	17
2.4	ソーシャル・デジタルライブラリにおける人物推薦	21
2.5	本論文の位置づけ	22
第3章	NDC ツリープロファイリング	25
3.1	ユーザプロファイリングに必要な特徴データの選定	25
3.2	NDC とは	29
3.3	NDC ツリープロファイリング	31
第4章	人物推薦アルゴリズム	35

4.1	NDC ツリープロファイリングを用いた人物推薦手法	35
4.2	推薦スコア算出式	38
4.2.1	推薦スコア算出式の提案	38
4.2.2	有効性の評価	40
4.2.3	実験データの集計	43
4.2.4	評価結果	45
4.3	推薦スコア算出式の改良	46
4.3.1	改良した算出式	46
4.3.2	有効性の評価	47
4.3.3	比較対象の推薦アルゴリズム	49
4.3.4	実験データの集計	51
4.3.5	評価結果	52
第 5 章	考察	57
5.1	NDC ツリープロファイリングの有効性	57
5.2	主題の粒度の影響	57
5.3	実験協力者ごとの差異	60
5.4	NDCTREE2 と BOOK の比較	66
5.5	NDC ツリープロファイリングの課題	69
5.5.1	実験条件の課題	69
5.5.2	類似尺度の妥当性	70
第 6 章	結論	73
	謝辞	75

参考文献	77
全研究業績のリスト	81

目次

1.1	本論文の主要構成	4
2.1	Ann Arbor District Library が提供する SOPAC の検索画面	8
2.2	BookSpace の画面イメージ	9
2.3	ソーシャルカタログギング・サイト ‘LibraryThing’ の画面イメージ	10
2.4	読書系 SNS‘ブックログ’ の画面イメージ	12
2.5	読書系 SNS‘読書メーター’ の画面イメージ	12
2.6	読書メーターの類似読者推薦機能	14
2.7	Henri によるオンラインコミュニティの発展段階	15
3.1	資料の分類に基づく階層構造として捉えたユーザの関心	29
3.2	NDC による日本の近代小説の分類 (階層表示形式)	30
3.3	NDC ツリーの例	32
4.1	評価 1: 実験システム ‘Shizuku2.0’ の操作画面	42
4.2	評価 1: Shizuku2.0 に登録されたコメントの件数 (ユーザ別)	45
4.3	評価 2: MyBookList における資料検索・登録画面	50
4.4	評価 2: MyBookList における資料利用履歴表示画面	50

4.5	評価 2: NDCTREE2 のユーザ別の F 値	54
4.6	評価 2: パラメータに対する F 値の平均値の推移	55
5.1	913.6 を除外した場合のパラメータに対する F 値の推移	59
5.2	実験協力者 14 番のパラメータに対する F 値の推移	62
5.3	実験協力者 14 番と推薦された実験協力者の NDC ツリー	64
5.4	実験協力者 11 番と推薦された実験協力者の NDC ツリー	65
5.5	実験協力者 11 番における, パラメータに対する F 値の推移	66
5.6	実験協力者 19 番と, 推薦された実験協力者の NDC ツリー	68
5.7	実験協力者 11 番が選好し, BOOK のみが推薦した NDC ツリー	70

表目次

4.1	評価 1: 実験協力者が登録した書籍の内訳	43
4.2	評価 1: NDC 第 1 次区分ごとの書籍登録冊数と割合	44
4.3	評価 1: 正解データの集計	44
4.4	評価 1: 各アルゴリズムに対する精度の平均と再現率の平均	45
4.5	評価 1: NDCTREE1 の精度, 再現率が最大の時のパラメータ	46
4.6	評価 2: 実験協力者が登録した書籍の内訳	51
4.7	評価 2: NDC 第 1 次区分ごとの書籍登録冊数と割合	52
4.8	評価 2: 正解データの集計	52
4.9	評価 2: 各アルゴリズムに対する精度の平均と再現率の平均	53
4.10	評価 2: NDCTREE2 の評価結果上位 11 件のパラメータ	53
5.1	小説を除外することによる精度・再現率・F 値の差異	58
5.2	実験協力者ごとの NDCTREE2 の精度・再現率・F 値	61
5.3	実験協力者ごとのアルゴリズムの比較	67

第1章

はじめに

1.1 デジタルライブラリの社会的文脈

デジタルライブラリは、World Wide Web(以下、ウェブ)の発達に伴い、紙メディア、電子メディアを問わず、情報資源にアクセスするシステムとして重要な役割を担ってきた。当初、デジタルライブラリの中心となる機能は、コンテンツの検索・閲覧機能であった。しかし、やがてデジタルライブラリが果たすユーザの社会関係への作用が、図書館情報学者に重視されているようになった。とりわけ、デジタルライブラリにおいて収集・保存されたデジタルコレクションは、利用者コミュニティが資料を利用する社会的な文脈が反映されているという指摘があり [1]、資料を仲立ちとした利用者コミュニティに対する社会的相互作用の支援への期待が高まっている [2]。このような背景から、デジタルライブラリの機能を定義した参照モデルには、デジタルライブラリのユーザ同士で、互いにコミュニケーションをとる機能が含まれている [3][4]。

デジタルライブラリにおけるユーザ同士の社会的相互作用は、デジタルライブラリの開発・運用の現場でも注目されている。Casey ら [5] は、管理者ではなく、ユーザが中心となって積極的にサービスを形成していくデジタルライブラリの構想として “Library2.0”

を提唱している。Library2.0は、図書館情報システムやOPAC(Online Public Access Catalog)に、SNS(Social Network Site)の機能や書評投稿機能を実装することで、構想の一部が実現されている。デジタルライブラリのユーザもまた、ユーザ同士の社会的相互作用に期待している。Brangier[6]は、フォーカスグループ・インタビューを通して、デジタルライブラリの基本的な機能要件を定義した。そこでは、同じ資料にアクセスしたユーザ同士で交流するための機能が含まれている。

このように、研究者およびデジタルライブラリの開発・運用関係者の両方から、デジタルライブラリの社会的相互作用の側面に注目が集まっている。

1.2 ソーシャル・デジタルライブラリ

Worrallは、資料を仲立ちとして社会的相互作用を生み出すデジタルライブラリのモデルとして、ソーシャル・デジタルライブラリを提唱している[7]。Worrallは、ソーシャル・デジタルライブラリにおける、コミュニティ形成と協働を実現する重要な概念として、Starら[8]が提唱しているバウンダリオブジェクトに着目している。バウンダリオブジェクトとは、複数のコミュニティの間の境界線を横断し、個々のコミュニティで独自の意味や構造が作り出される用語、組織、物体などを指す用語である。このような性質によって、バウンダリオブジェクトは、コミュニティ間でその意味や構造について議論する機会を生み出し、各コミュニティの規範や思想を明らかにすると共に、コミュニティ間の共通項を見出させ、新たなコミュニティ形成の媒介となる。Worrallは、資料がソーシャル・デジタルライブラリのバウンダリオブジェクトとしての役割をはたし、新たなコミュニティ形成を促進すると主張している。このような仮定に基づき、Worrallは、類似した環境下におけるユーザの行動についての調査を行った。その結果、ソーシャル・デジタルライブラリがユーザ間で関心を共有させ、議論を活発化させ、様々な相互作用を生むこと

を明らかにした [7]. このことから、デジタルライブラリにおける社会的相互作用の機能は重要であり、具体的な実装の必要性が導かれている。ただし、バウンダリオブジェクトとして機能する要素が具体的に何か、その単位やユーザの関心との関係などについて、Worrall は明確に定義していない。また、バウンダリオブジェクトを通じたコミュニティ形成支援を、システムの機能として提供するための具体的な提案については未だ行われていない。以上から、ソーシャル・デジタルライブラリのコミュニティを強化するためには、バウンダリオブジェクトの提供方法と、それを活用してデジタルライブラリ内でコミュニティ形成を支援する機能の実装手法を提案する必要がある。本論文ではこの課題を解決するために、ソーシャル・デジタルライブラリ上に、コミュニティ形成支援のための機会を作るための人物推薦手法を提案する。

1.3 本論文の構成

本論文の構成を図 1.1 に示す。第 1 章ではここまで、本研究の研究背景と研究目的を述べた。そのうえで、ユーザ間で社会的相互作用を促進するデジタルライブラリとして、ソーシャル・デジタルライブラリの研究および開発が進んでいるが、ソーシャル・デジタルライブラリの重要な概念であるバウンダリオブジェクトを提供する手法は未だ明確ではないことを指摘した。そこで本論文では、ソーシャル・デジタルライブラリにおける人物推薦手法を提案することを目的とする。

第 2 章では、関連研究について記述する。まず、本章で概念の定義などを記述したソーシャル・デジタルライブラリの現状について説明する。ソーシャル・デジタルライブラリは、ソーシャル OPAC とソーシャルカタログ・サイト、読書系 SNS の 3 種類のシステムとして実現化されていることを示す。そのうえで、ソーシャル・デジタルライブラリの現実の取り組みの中で、ユーザ間の相互作用やコミュニティ形成に影響を与える重

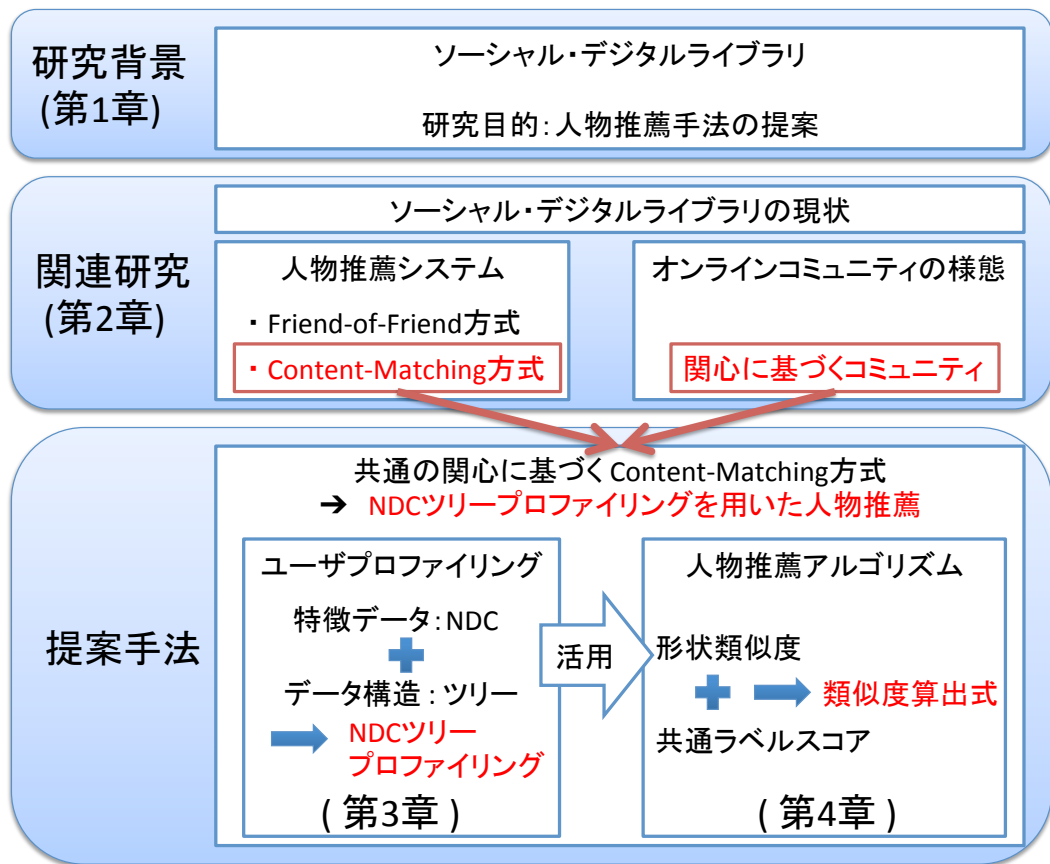


図 1.1 本論文の主要構成

要な機能として、類似読者推薦機能に着目し、ユーザ間の類似度を算出する尺度についての検証がなされていないことを指摘する。次に、オンラインコミュニティと学習の関わりについての関連研究を概説し、さらに、人物推薦システムの定義および関連研究について概説し、ソーシャル・デジタルライブラリにおける人物推薦のアプローチとして、Content-Matching方式の人物推薦手法が望ましいことを主張する。そのために、ユーザの関心を表現するユーザプロファイリング手法と、ユーザ間の関心の類似度を算出する人物推薦アルゴリズムの提案を、本論文の課題として設定する。

第3章では、ソーシャル・デジタルライブラリにおけるユーザプロファイリング手法を検討する。ユーザプロファイリングを実現するためには、プロフィールの構成要素となる特徴データを収集する必要がある。特徴データには、ユーザの属性に関するデータと、

ユーザの関心を表わすデータの2種類があるが、ここではユーザの関心を表わすデータである必要がある。ソーシャル・デジタルライブラリにおいて、ユーザに関するデータとして代表的なものに、ユーザの読書記録や貸出記録などの資料利用履歴がある。資料利用履歴には、資料の主題を表わす分類番号が含まれており、分類番号を特徴データとして抽出することで、ユーザの関心の傾向を大枠で推測することが可能である。このような検討に基づいて、ユーザの資料に対する関心の傾向を表わすために、資料利用履歴に含まれる分類番号として、日本十進分類法 (NDC) 番号を利用することを提案する。具体的な利用方法は、NDC 番号を、その階層性を利用して、ツリー状のデータとしてユーザプロフィールを作成する。本論文では、このツリー状のデータを NDC ツリーと呼び、NDC ツリーを作成する一連の処理を NDC ツリープロフィールと呼ぶ。

第4章では、第3章で提案する NDC ツリープロフィールに基づく人物推薦アルゴリズムの提案と評価を行う。人物推薦アルゴリズムには、ユーザプロフィール同士の類似度を測る尺度が必要である。本論文では、「形状類似スコア」「共通ラベルスコア」の2つの尺度を提案する。形状類似スコアは、NDC ツリーがツリー構造である特徴を利用し、ツリーの編集距離に基いてプロフィール間の類似度を計算する尺度である。形状類似スコアは、NDC ツリー間の差異に着目して類似度を計算するため、ユーザ間で共通していない関心が多ければ多いほど、形状類似スコアは低下する。共通ラベルスコアは、NDC ツリーに付与されている NDC 番号のラベルを比較し、共通するラベルの件数を NDC ツリーの階層毎に点数化し、集計したスコアである。本論文では上述の2つの類似度に加えて、「コメント率」という補助的な尺度を導入する。コメント率は、ソーシャル・デジタルライブラリにユーザが投稿した、資料に関するコメントの比率を算出した尺度である。本論文では、以上の3種類の尺度を組み合わせた推薦スコア算出式を提案し、実験により精度・再現率を評価する。そのうえで、評価結果の知見に基づいて、推薦スコア算出式を

改善し、最終的な人物推薦アルゴリズムを、本論文の手法のひとつとして提案する。

第5章では、第4章の実験結果に基づいて、NDC ツリープロファイリングによる人物推薦の特徴および課題について議論・考察する。

第6章では、第2～5章までの総括を記述したうえで、第5章の考察を踏まえた、ソーシャル・デジタルライブラリにおける人物推薦手法の提案をまとめる。

第2章

関連研究

2.1 ソーシャル・デジタルライブラリの現状と課題

既存のソーシャル・デジタルライブラリは、ソーシャル OPAC、ソーシャルカタログ・サイト、読書系 SNS という3つの関連領域の中で別々に運用・推進されている。

1 番目のソーシャル OPAC は、Library2.0 というコンセプトの実現を目指して開発された OPAC の総称である [9]。Library2.0 とは、図書館員である Michael Casey が提唱しているコンセプトであり、図書館サービスの創造と評価にユーザが関わることを促すための、ユーザ中心の図書館モデルと定義されている [5]。ソーシャル OPAC の開発はこの Library2.0 をデジタルサービスの側面から実現することを志向した試みである。ソーシャル OPAC の主な機能としては、ユーザによるタグ付け機能や書評の投稿機能、ソーシャル・ネットワーク・サイト (SNS) といった機能が挙げられる。ソーシャル OPAC の代表的な事例として、SOPAC(<http://thesocialopac.net/>) と BookSpace(<http://www.hclib.org/pub/bookspace>) がある。SOPAC は、アメリカ合衆国の州立図書館である Ann Arbor District Library が主導して開発したソーシャル・デジタルライブラリであり、SNS や書評投稿などの機能を提供する。図 2.1 は、Ann Arbor

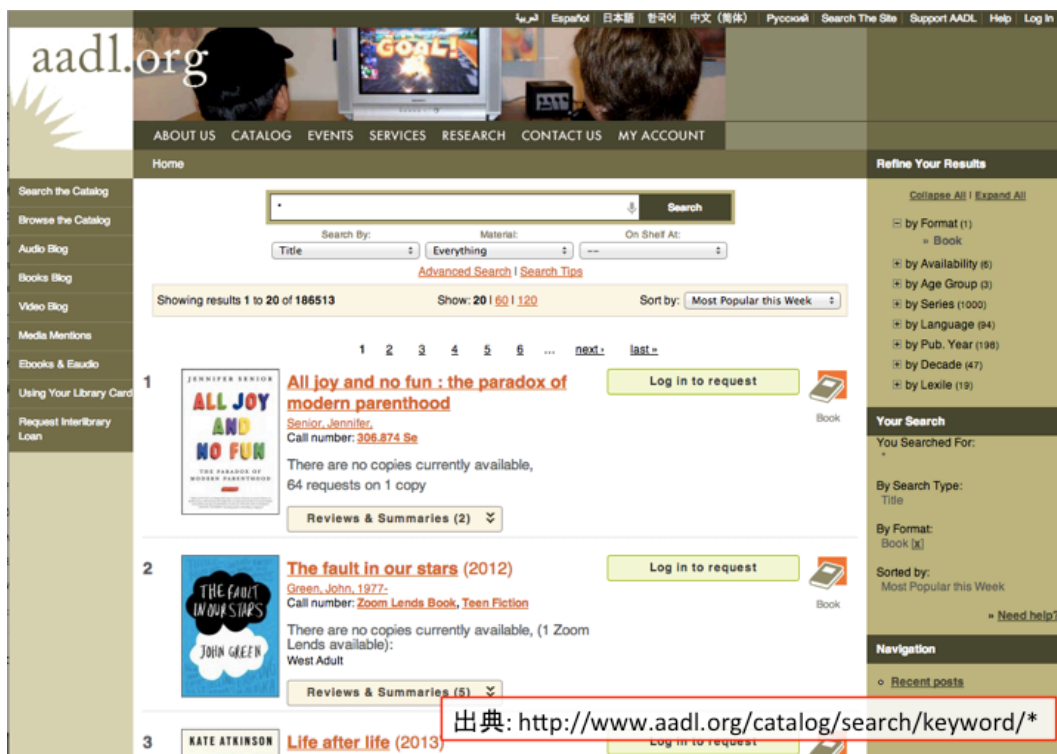


図 2.1 Ann Arbor District Library が提供する SOPAC の検索画面

District Library が提供しているソーシャル OPAC の検索画面である。もうひとつの BookSpace は、アメリカ合衆国の公立図書館である、Hennepin Country Library(HCL) の OPAC に実装されたサービスである (図 2.2)。ユーザは、従来の OPAC の機能に加え、それぞれが好む資料のリストを“Book List”として投稿することができる。そして、それぞれの Book List に対して、他のユーザが自由にコメントを付与することで、ユーザ同士がコミュニケーションを交わすことができる。日本においても、九州大学が既存の地域 SNS と連携して OPAC にコミュニティ機能を持たせた例がある [10]。この他にも、オープンソースの OPAC である VuFind が資料の検索結果からユーザ独自のリストを作成、保存する機能を提供するなど、他の図書館が導入可能なソーシャル OPAC のソフトウェアも開発されている。

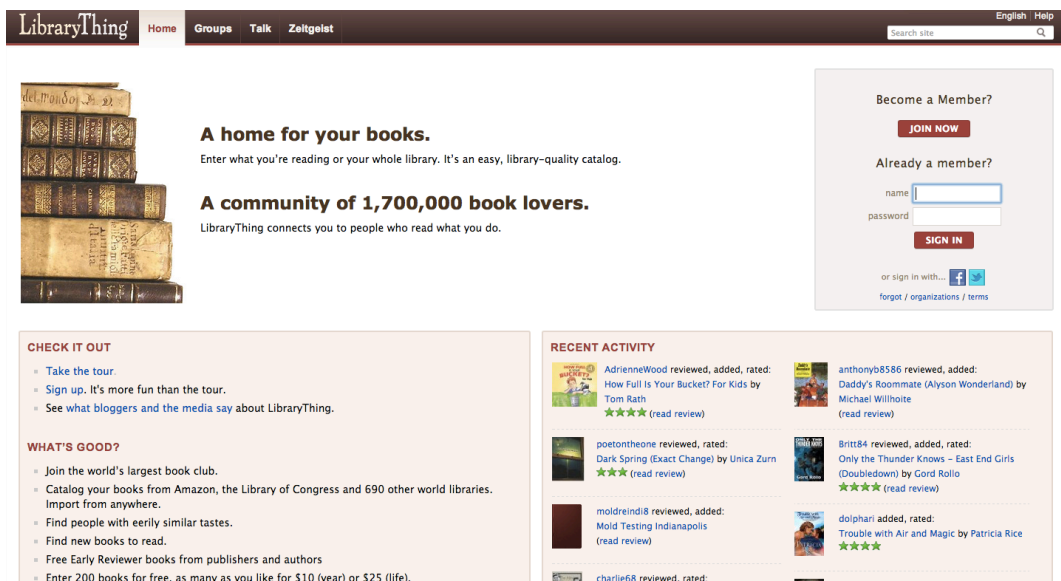
2 番目のソーシャルカタログ・サイトは、ユーザの協働によって目録を作成する取り組みやウェブサービスを指す。従来の図書館蔵書目録は、新たな書誌の追加や項目



(出典: <http://www.hclib.org/pub/bookspace/profile.cfm>)

図 2.2 BookSpace の画面イメージ

などの編集権限は、図書館関係者に限定されていた。そのため、OPAC のユーザは、ある資料について他のユーザに有益な情報を知っていたとしても、その情報を目録に反映することはできなかった。また、目録情報に誤りや項目の欠損があったとしても、ユーザは図書館員または OPAC の管理者に改善を申請するしか手段がなかった。ソーシャルカタログ・サイトは、ユーザが目録の質の向上に寄与する機会を損失しているという問題を解決するため、ソーシャル・ブックマークサービスや Wiki などを参考にして構築されたウェブサービスの総称である。代表的なソーシャルカタログ・サイトの事例として LibraryThing(<http://www.librarything.com>, 図 2.3) が挙げられる [11]。LibraryThing では、アカウントを作成したユーザが、自身の仮想的な本棚をサービス上に作成することができる。LibraryThing には、あらかじめ既存の図書館目録を取り込ん



(出典: <http://www.librarything.com>)

図 2.3 ソーシャルカタログ・サイト ‘LibraryThing’ の画面イメージ

であり、ユーザは本棚に登録する資料の情報をすべて登録すること無く、一般に流通されている出版物を検索し、本棚に資料を登録することが可能である。しかし、LibraryThing にユーザが登録したい資料が未だ登録されていない場合には、ユーザ自身が書誌情報や所在情報を登録することができる機能が提供されている。このように、ユーザひとりひとりの貢献を集約することで、これまでに全世界で刊行された出版物の目録を作成することを目指すのが、LibraryThing をはじめとするソーシャルカタログ・サイトの目的である。さらに、ソーシャルカタログ・サイトでは、目録の協働作成の機能を通して、ユーザ間のコミュニケーションを促進する作用もある [12]。ソーシャル・デジタルライブラリの提唱者である Worrall は、このような作用に着目して、LibraryThing に対するユーザの行動調査を行った [7]。

最後に読書系 SNS であるが、これは、本を媒介として読書家のソーシャル・ネットワークを作り出すウェブサービスである。増井 [13] は 2005 年に、「ユーザが本棚形式で蔵書管理を行いつつ、他人の本棚を覗いたりコメントを共有したりすることによって、緩やか

なグループコミュニケーションを行なうことが可能」とすることを旨としたウェブサービスである本棚.orgを開発した。本棚.org以降、ブックログ (<http://booklog.jp>, 図 2.4) や読書メーター (<http://book.akahoshitakuya.com>, 図 2.5) など、多くの類似サービスが開発・運営されるようになった。読書系 SNS で共通する機能としては、本棚作成機能とソーシャル・ネットワーキング機能、そして掲示板機能が挙げられる。本棚作成機能とは、ユーザがウェブサービス上に本のリストを仮想上の本棚として登録し、他のユーザに公開する機能である。読書メーターやブックログなどの主要な読書系 SNS では、本棚に登録した本それぞれに対して、「読みたい」「読書中」「読了」などの読書状況を登録する機能も提供されており、本棚作成機能は実質的にユーザの読書記録として活用できる。ソーシャル・ネットワーキング機能は、ユーザ同士が公開された本棚の閲覧機能やユーザの検索機能などの提供を通して、ユーザがコミュニケーションをしたいと考える他のユーザを発見することを促し、ユーザ同士が友達として関係を明示的に登録することを可能にする機能である。掲示板機能は、ユーザが読書に関するテーマで自由にスレッドを立ててコミュニケーションをとれるような電子掲示板機能である。このように、ユーザが本棚という形式で読書記録を他者と共有し、そのうえでコミュニケーションを行い読書に関するソーシャル・ネットワークを構築するのが読書系 SNS の目的である。ソーシャルカタログ・サイトと読書系 SNS は機能に重複がみられる。ただし、読書系 SNS の多くは、本棚機能で参照される書誌情報は Amazon などの外部のオンライン書店のデータに依存しており、ソーシャルカタログ・サイトが提供するような目録を補完する機能は一部を除き提供されていない。また、読書系 SNS は書籍の所蔵情報を持たないという点で、ソーシャル OPAC と異なる。

こうした3つの関連領域のサービスは、機能の差異がみられるが、ユーザが特定の情報資源を媒介としてコミュニケーションを行うことを支援するシステムであるという点で共



(出典: <http://booklog.jp/users/kunimiya>)

図 2.4 読書系 SNS‘ブックログ’の画面イメージ



(出典: <http://book.akahoshitakuya.com/>)

図 2.5 読書系 SNS‘読書メーター’の画面イメージ

通している。その中でも読書系 SNS は、ソーシャル・デジタルライブラリが目指す資料を仲立ちとしたコミュニケーションの実現という点で、ソーシャル・デジタルライブラリに最も関連が強い。

ソーシャル・デジタルライブラリの課題として、第 1 章で挙げたユーザ同士の交流を直接支援する機能については、ソーシャルカタログ・サイトと読書系 SNS において、類似読者推薦機能が対応する機能として提供されている。類似読者推薦機能は、ユーザが登録した読書記録に基づき、ユーザにより近い関心を持つユーザを検索し、コミュニケーションの相手として推薦する機能である。読書系 SNS である読書メーターでは、図 2.6 のように、類似読者推薦機能を「相性のよい読書家さん」という名称で提供している。この機能は、ユーザが一定量以上の本の感想をサービス上に投稿すると、投稿された本と付随する感想文を分析し、類似した読書傾向を持つとサービスが推測したユーザをランキング形式で提示する。このような機能を提供することで、読書メーターではユーザがコミュニケーションを取りやすいような相手を発見することを支援し、サービス内のソーシャル・ネットワークを豊かにすることを促している。類似読者推薦機能は、読者のソーシャル・ネットワークを作成するという読書系 SNS の目的に直接関係することから、ソーシャル・デジタルライブラリにおいて重要な機能であるといえる。ただし、類似読者推薦機能が基準とする読書傾向の類似性の尺度の妥当性については、明確に研究が為されていない。ユーザ間の読書傾向の類似性は、登録された資料のレベルでの一致度や、主題の関心のようなより抽象的な一致度など、様々な粒度で考えることが可能である。ソーシャル・デジタルライブラリにおいて、読書傾向の類似性の尺度の妥当性を検討することは、ユーザ間の社会的相互作用を引き起こすバウンダリオブジェクトは何かという、本質的な命題につながる試みである。

以降の節では、オンライン上のユーザ間の社会関係に関する研究領域であるオンライン

The screenshot shows the 'Read Meter' website interface. At the top, there are navigation links for 'Home', 'My Page', 'Common Reading', 'Thoughts on Favorites', 'Compatibility', and 'Other'. A search bar is present with the text '本のタイトル・著者名を入力' (Enter book title or author name). Below the search bar, there is a section titled '相性のよい読書家さん' (Readers you are compatible with). A warning message states: '読んだ本を登録したり、感想・レビューを投稿するとさらに相性占いの精度が上がります。気になる読書さんがいましたらお気に入り登録してみましょう！また、相性データの更新には1〜3時間ほどかかりますので、本を登録したら時間をおいて確認してみてください。' (Registering books you've read or posting thoughts/reviews will further improve the accuracy of the compatibility占卜. If you're interested in a reader, please register them as a favorite! Also, updating compatibility data takes about 1-3 hours, so please check back after some time after registering books.)

The first recommendation is for the reader '備忘録' (Memoirs) by Naho Sotome, with a compatibility score of 86%. Below this are four book recommendations:

Book Title	Author	Registration Count
クビキリサイクル 青色サヴァンと戯言遣い	西尾 維新	4393登録
死神の精度	伊坂 幸太郎	5429登録
颯麿の匣 (講談社ノベルス)	京極 夏彦	2462登録
ヒストリエ(8) (アフタヌーンKC)	岩明 均	829登録

(出典: <http://book.akahoshitakuya.com/aishou>)

図 2.6 読書メーターの類似読者推薦機能

コミュニティと人物推薦システムを関連研究として記述する。それによって、類似読者推薦機能に適した、読書傾向の類似尺度についての検討を具体的な課題として明確化する。

2.2 オンラインコミュニティの様態

ソーシャル・デジタルライブラリのユーザが形成したいオンラインコミュニティは多様であり、ソーシャル・デジタルライブラリは、ユーザそれぞれの欲求に対応したコミュニティ形成支援機能を提供する必要がある。しかし、ユーザの欲求を個別に満たすには限界がある。なぜなら、コミュニティはユーザ間の関係性が基盤となって形成されるため、実際に形成できるコミュニティは、ソーシャル・デジタルライブラリのユーザ群の属性や関

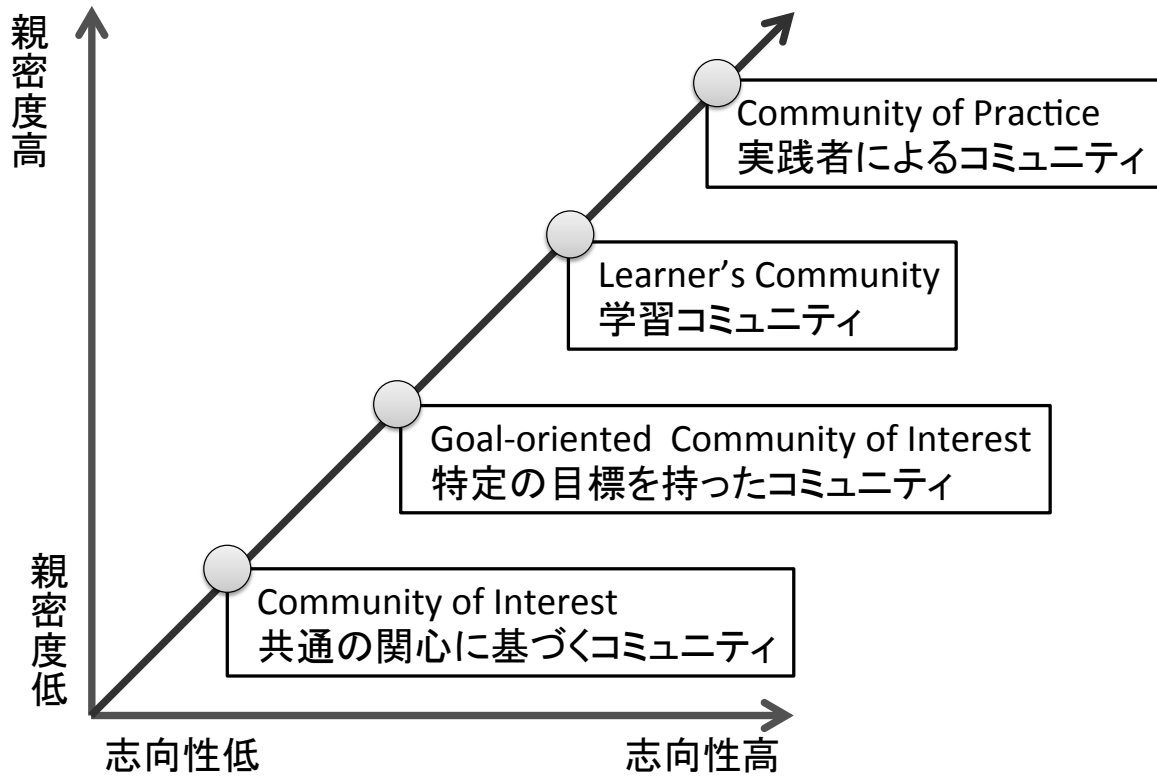


図 2.7 Henri によるオンラインコミュニティの発展段階

心によって制限されるためである。そのため、コミュニティ形成の支援を実現するためには、オンラインコミュニティに関する先行研究の知見に基づいて、多様な利用者のニーズを包含するような、共通のコミュニティ形成の要件を見出し、その要件を満たす基盤を作る必要がある。

Henri ら [14] は、図 2.7 のように、学習に関わるオンラインコミュニティの関連概念を整理し、「関心コミュニティ」「目的志向型の関心コミュニティ」「学習者コミュニティ」「実践コミュニティ」の 4 つに区分している。さらに、これらのコミュニティ概念の関係を、構成員間の目的意識と結束の強さにしたがって発展していくコミュニティのプロセスとして位置付けている。関心コミュニティは、共通の関心に基づいて集まった人々によって構成されたコミュニティであり、Fischer[15] によって定義された。関心コミュニティは

本来別の実践コミュニティに所属する人間同士によって形成されるコミュニティであり、特定の課題や目的を持たず、主にコミュニティを構成する共通の関心に関する情報を交換する。目的志向型の関心コミュニティは、関心コミュニティがコミュニケーションの過程で共通の目的意識を持つことで発展するコミュニティである。目的志向型の関心コミュニティでは、情報の交換がある共通の目的を達成するために行われ、このタイプのコミュニティでは最終的な成果物を形作ることが前提となる。学習者コミュニティは、目的志向型の関心コミュニティの過程で結束を強めた構成員が、継続的な情報収集と学習ノウハウを交換する場として形成されるものである。実践コミュニティは、ある目的を志向し、実践を通して知識の創出や活動を行うコミュニティであり、Wengerら[16]が定義した。実践コミュニティの理論では、コミュニティが抱える課題の解決に向けた実践に構成員が関与すると同時に、コミュニティにおける実践の意義を明示化することを学習と定義している。実践コミュニティの理論は、ある固有の知識を入力するという従来の学習観とは異なり、コミュニティでの構成員間の実践による相互作用が、学習に本質的に必要な行為であることを主張している。このように、様々なオンラインコミュニティは、関心コミュニティを起点として段階的に発展していく過程で形成されるとHenriは主張している。したがって、関心コミュニティの形成を支援することは、他のタイプのコミュニティが成り立つためのための基盤を形成することに繋がる。そのため、ソーシャル・デジタルライブラリにおける、コミュニティ形成支援の基本要件のひとつとして、関心コミュニティの形成を目的とした機能の提供が挙げられる。

関心コミュニティを形成するには、ユーザ間の知識体系の仲介物を提供することが必要である。Fischer[15]は、関心コミュニティを形成する際に、それぞれの構成員がもともと所属している、実践コミュニティの知識体系に対する食い違いが、共通の関心を認識することを困難にさせているという障害があると主張している。そのうえで、これを解決する

手段として、構成員間の知識体系を仲介するような、共通の知識体系を表現するオブジェクトを提供するシステムを提案している。

本論文では、ソーシャル・デジタルライブラリにおけるコミュニティ形成支援として、関心コミュニティの形成を志向する。そのうえで、コミュニティを形成するアプローチとして、人物推薦システムという領域に注目する。

2.3 人物推薦システム

人物推薦システムの研究は、推薦システムの研究と専門家検索の研究の2つの潮流から合流してできた研究分野である。推薦システムとは、ユーザが要求する情報をシステムが能動的に推測し、提示するシステムのことを指す。初期の推薦システムは、内容ベースフィルタリングを採用することが多かった。これは、メールなどのテキストベースのコンテンツを分析し、ユーザの嗜好を抽出したうえで、コンテンツをフィルタリングして提示するというアプローチである。また、協調フィルタリングと呼ばれるアプローチも広く用いられてきた。協調フィルタリングは、複数のユーザの行動履歴を分析し、類似したユーザのコンテンツ利用履歴から、ユーザの好むコンテンツを推測するというアプローチである。推薦システムは、ユーザが明確に意識していない情報欲求に対しても、それに見合ったコンテンツを提示することができるという点で、能動的にクエリを入力し検索を行う必要のある情報検索の技術と比べて、ユーザの負担が少ないといった利点がある。

専門家検索の研究者は、このような推薦システムの利点を、情報の取得だけに留まるのではなく、ユーザの社会関係に関する欲求を満たすように応用した。専門家検索とは、ユーザが専門家からの助言を必要とする際に、協力を要請する専門家を、人物データベースから検索するための技術を指す。初期の専門家検索は、ユーザが関心を持つ専門分野の用語等をクエリとしてシステムに入力する必要があるため、ユーザに一定の知識が要求さ

れるという課題があった。この課題を解決するために、協調フィルタリングを応用した‘Referrall Web’が開発された [17]。Referrall Web は、論文の共著関係などの専門家のソーシャル・ネットワークを分析して、ユーザに関連のある専門家を、推薦する専門家検索システムである。また、内容ベースフィルタリングの推薦システムの応用も行われており、そこでは、ユーザの投稿した記事コンテンツの内容を分析することで、ユーザの分野と類似した専門家を検索するシステムを開発した。

初期の人物推薦システムは、推薦システムの応用事例として研究されてきた。しかし、その後、社会関係を構築することを支援するという独自性に着目し、人物推薦システムは特有の問題を扱った独自領域として定義されるようになった。Terveen ら [18] は、社会関係の構築を支援するための人物推薦システムや専門家検索システムなどを総称してソーシャルマッチングと呼んでいる。ソーシャルマッチングでは、単純に人物を発見するだけでなく、発見した人物をどのようなインターフェースで提示するか、あるいはユーザ同士が実際に社会関係を構築するためのコミュニケーションを支援する場をいかに提供するか、そして成立した社会関係を次の人物推薦の精度向上にどのようにフィードバックするか、といった複数の過程における支援として何が必要かが重要な課題であると、Terveen らは主張している。このような、社会関係を構築する環境全体を意識した研究の1つとして人物推薦システムが位置づけられるようになっている。

人物推薦システムは、ウェブにおける SNS の登場によって、日常の場面で求められるようになった。SNS とは、人物間の社会関係を可視化するためのウェブサイトである [19]。ユーザは SNS 上で、社会関係のある他のユーザを検索し、つながりのある人物としてサイトに登録することができる。さらにユーザは、他のユーザの社会関係を表すリストを参照することができ、そのようなリストを辿ることで、さらにソーシャル・ネットワークの可視化を進めることができる。また、ユーザ間でコミュニケーションのツールが提供

されることも多い。現在では、代表的な SNS として Facebook や Twitter などが世界的に普及しており、ユーザの社会関係を活用するウェブサービスの基盤となるなど、現在のウェブを支える基盤の 1 つとして機能している。このような SNS において、既に構築している社会関係を再現するために、ユーザにとって既知のユーザを推測し、友達として登録することを促すための機能として、人物推薦が活用されている。実際に、Facebook や Twitter などの代表的な SNS では、人物推薦システムが機能の 1 つとして実装されている。

また、ウェブ上で既に存在する社会関係だけでなく、オンライン上で新たな人物と出会い、社会関係を新たに形作るためのサービスも発展している。具体的には、結婚相手などのパートナーを探すためのウェブサイトや、共通の関心を持つ人物同士の同好会コミュニティのウェブサイトなどが挙げられる。こうした新規に社会関係を構築することを目的としたウェブサイトはマッチングサイトと呼ばれており、前述したソーシャルマッチングの研究が応用されている。マッチングサイトにおける人物推薦では、ユーザの交友関係の嗜好を推測することが必要である。ユーザの交友関係に関する嗜好は大別して、相手の所属や年齢、性別などのユーザの属性に関する嗜好と、趣味や好きな作家、関心のある学術領域などのユーザの関心に関わる嗜好の 2 種類が存在する。マッチングサイトにおける人物推薦システムは、ユーザの属性と、ユーザの関心のどちらに対する嗜好に基づいて構築されることを望むのかを検討した上で、それに適したユーザの嗜好の推測と、その嗜好にあてはまる他のユーザを発見する手法を検討する必要がある。

人物推薦を実現するには、ユーザが望む社会関係を検討した上で、それを実現するためのユーザの選好を分析する技術と、推薦アルゴリズムの提案が必要である。一般に、人物推薦のアプローチには、Content-Matching 方式と Friend-of-Friend 方式の 2 種類がある [20]。Content-Matching 方式は、人物が投稿したコンテンツの内容を分析すること

によって、ユーザが好む人物を推測するアプローチである。対して、Friend-of-Friend 方式では、Facebook における友達リストや、Twitter におけるフォロワー・フォロー関係など、ユーザが既に SNS に登録しているソーシャル・ネットワーク情報を分析し、ユーザと関連性が高いと考えられる人物を推測するアプローチである。Content-Matching 方式は、ユーザの関心に基づいて共通の関心を持つユーザを発見するのに適している。一方、Friend-of-Friend 方式はユーザの関心とは異なり、ユーザの交友関係のパターンから望ましい交友関係を築く可能性の高いユーザの発見に適している。こうした傾向から、Friend-of-Friend 方式は、交友関係の構築のための人物推薦に用いられることが多い。反対に Content-Matching 方式は、共通の関心に関するコミュニティを構築する場合に用いられる。ソーシャル・デジタルライブラリにおける人物推薦は、ユーザ間で共通する関心を見出すことが重要であるため、Content-Matching 方式のアプローチを選択することが望ましい。

Content-Matching 方式に基づいた人物推薦を実現するためには、次のような課題を達成する必要がある。まず人物推薦全般の課題と、ソーシャル・デジタルライブラリ特有の課題の2つの課題に対して、それぞれのアプローチを決定する必要がある。人物推薦全般の課題として挙げられる問題として、ユーザがどのような人物を好むか、そしてユーザ間の関係がどのようなものであることが望ましいのかという概念的な2つの課題と、推薦を行うためにユーザの関心をいかに推測し、ユーザ同士の関心の類似度を算出するかという、実装上の2つの課題について検討する必要がある。そして、ソーシャル・デジタルライブラリにおける人物推薦に特有の課題として、ユーザの関心をどのように抽出、推測するかということ、そして、ユーザ間の関心の類似度をいかに算出し、スコアとして反映させるかを検討する必要がある。

2.4 ソーシャル・デジタルライブラリにおける人物推薦

ソーシャル・ネットワークにおける人物推薦の目的は、ユーザ間で社会関係を構築し、コミュニケーションを誘発するための仕組みを作ることである。その社会関係やコミュニケーションの様態は人物推薦を適用する領域によって変わる。本論文では、社会関係の様態によって人物推薦を3種類に大別する。

1. 専門家からの助言を得るための人物推薦
2. 交友関係を構築するための人物推薦
3. 共通の関心コミュニティを構築するための人物推薦

専門家からの助言を得るための関係とは、専門家検索という領域で研究されてきた内容である。ここでは推薦対象となる人物と推薦結果を受け取るユーザは非対称の関係性を持っており、推薦対象となるユーザは、何らかの点でユーザを上回る知識を持っているということが前提である。基準として、ユーザの情報欲求に適した情報を有する人物を探索するということが、専門家検索における人物推薦の課題となる。デジタルライブラリの領域においては、こうした関係性は、ライブラリアンとユーザの間でチャットやメールなどを用いて行う情報案内サービスである、デジタル・レファレンスサービスの場面で適用される人物推薦の様態である。

交友関係を構築するための人物推薦とは、出会い系サイトのような、何らかのパートナーを得るためのコミュニケーションを支援するための人物推薦である。Facebook や Mixi などのようなソーシャル・ネットワーク・サイトにおいては、既存の友人関係をオンライン上で再現するために適用されることもある。この種の人物推薦で重要な点は、この関係性は相互性を有しているという点である。相互性とは、推薦対象と推薦結果を受ける

ユーザ双方の嗜好が、推薦結果の成否に影響を与える性質を指す [21]. 出会い系サイトのような場面では、互いに好感を抱いていない場合には、たとえ片方にとって好ましいユーザとしてあるユーザが推薦されたとしても、本来の目的である社会関係の構築やコミュニケーションの誘発にはいたらない可能性が高い。相互性を有する人物推薦が成功するためには、相互の嗜好が調和するような組み合わせを考える必要がある。

共通の関心コミュニティを構築するための社会関係は、特定の主題に関して、他者との意見交換や議論を望むユーザを対象とした人物推薦である。関心コミュニティの事例としては、映画や小説などの作品・著者のファンコミュニティや、ある学術領域の学会コミュニティなどがあてはまる。こうした場面における、他者に対する選好基準は、共通した関心を持っているか否かにある。交友関係を構築するための人物推薦とは異なり、相互性は比較的強く、互いに共通のトピックを持っているか否かが重要である。

2.5 本論文の位置づけ

本論文は、ソーシャル・デジタルライブラリにおける人物推薦手法を提案することを目的とする。オンラインコミュニティなどの関連研究から、ソーシャル・デジタルライブラリにおけるコミュニティは、共通の関心を持つオンラインコミュニティであることが望ましい。そのうえで、本論文では関心コミュニティの形成を目的とした人物推薦のアプローチとして、Content-Matching 方式を選択した。このことから、共通の関心を持つユーザを推薦するための、Content-Matching 方式の人物推薦技術の提案が本研究の課題である。そのためには、ソーシャル・デジタルライブラリ上で実現可能なユーザプロファイリングと、人物推薦アルゴリズムを提案する必要がある。そこで本研究では、ソーシャル・デジタルライブラリにおけるユーザプロファイリングとして NDC ツリープロファイリングを、人物推薦アルゴリズムとして形状類似スコアと共通ラベルスコアを要素とする人物

推薦スコア算出アルゴリズムを提案することで課題を解決する。

本論文のアプローチはソーシャル・デジタルライブラリの研究において、次の3点の独自性を持つ。1点目は、ユーザ間の関心の類似を、資料の共通性ではなく主題の共通性で算出する点である。従来のソーシャル・デジタルライブラリの類似読者推薦機能では、主に資料を単位としてユーザの関心を推測するアプローチが取られている。しかし、資料は大量に存在するため、同じ関心を抱いていたとしてもその共通性を見出しにくい。本論文ではより共通性を見出しやすい範囲で、ユーザの関心同士の類似を見出す方針を採った点に独自性がある。

2点目は、ユーザが関心を持つ主題を表現するために、図書の分類規則を用いる点である。推薦システムの研究において、ユーザプロファイリングの主要なアプローチは、ユーザの選好または投稿したコンテンツからキーワードを抽出して、ベクトル空間モデルなどの特定のデータ構造に従ってプロファイリングを行う手法である。しかし、キーワードを用いる場合には、同義語や類義語などの判定などを行う必要がある。本論文では、ユーザの関心が資料利用履歴に反映されており、資料の主題は分類番号を抽出することで把握できるという点に着目し、ユーザプロファイリングに図書の分類規則を用いた点に独自性がある。

3点目は、ユーザプロファイルを階層構造にした点である。ユーザ同士で共有する関心の粒度は、ソーシャル・デジタルライブラリを利用しているユーザの規模や、関心の傾向によって異なる。本論文では、階層構造のプロファイルを採用することで、ユーザ同士の類似を測るための、共通する主題の粒度を状況に応じて変更することを可能にする。

第3章

NDC ツリープロファイリング

3.1 ユーザプロファイリングに必要な特徴データの選定

Content-Matching 方式の人物推薦におけるユーザプロファイリングは、ユーザの関心を推測するための手がかりとなる特徴データを収集し、特定のモデルに基づいてプロファイルを作成する。そのため、ソーシャル・デジタルライブラリにおけるユーザプロファイリング手法を検討するには、まずソーシャル・デジタルライブラリが記録しているユーザの行動データの中から、適切な特徴データを選定する必要がある。ソーシャル・デジタルライブラリにおいて、ユーザに関するデータとしては、資料利用履歴とユーザの投稿コンテンツの2種類が挙げられる。資料利用履歴は、ユーザがアクセスした資料を時系列に記録したものであり、通常の図書館では貸出利用履歴がこれに対応する。また、デジタル情報資源のみを提供するようなデジタルライブラリでは、ウェブを通じた資料へのアクセスログがこれに対応する。その他、ソーシャル・デジタルライブラリの中にはユーザが任意の資料を指定すると資料のリストを作成する機能があり、この資料のリストもまた資料利用履歴に準じるものとして活用できる。ユーザの投稿コンテンツは、ソーシャル・デジタルライブラリが提供する資料に付与する情報としてユーザが独自に作成したコンテンツを

指す。具体的には書評記事や、タグ付けなどのアノテーション、ユーザ同士のチャット・掲示板の投稿履歴などが挙げられる。資料利用履歴とユーザの投稿コンテンツはそれぞれユーザの行動履歴を表したデータであるといえる。このうち、ユーザの投稿コンテンツに関してはソーシャル・デジタルライブラリの機能に依存するデータであり、必ずしもシステムが自動的に入手できるとは限らない。これに対して、デジタルライブラリが資料を提供するシステムであるため、資料利用履歴は存在することが前提となっている。そこで、本研究ではソーシャル・デジタルライブラリのユーザの特徴データを、ユーザの資料利用履歴から抽出するアプローチを採用する。

資料利用履歴の主要な構成要素は、資料の書誌情報である。そのため、資料の書誌情報からいかにしてユーザの関心を把握するための特徴データを抽出するかを考える必要がある。特徴データの抽出について考える際に、データの粒度を考えることは、ユーザの関心を正確に把握するうえで重要である。ソーシャル・デジタルライブラリにおける既存の類似読者推薦機能は、資料を最小単位として、資料利用履歴に登録された資料の一致度によって、ユーザの関心の類似度を計算している。しかし、ユーザの関心を表すために資料を最小単位とすることが必ずしも適切とは限らない。ユーザ同士の関心の類似度を正確に把握することに失敗する場合が考えられる。一般的にデジタルライブラリにおける資料の登録件数は膨大であり、ユーザひとりが利用する資料の件数には限界がある。そのため、ユーザ同士の資料利用履歴を比較した場合に、共通する資料の点数が少ない傾向に陥る可能性がある。共通の資料が無くとも、例えば数学や物理学といった資料の主題がユーザ間で共通している場合、これらのユーザの関心は類似していることが十分考えられるが、資料を単位とすると、類似したユーザを発見することができないことになる。これは、本来コミュニケーションをとる相手として推薦すべきユーザを発見できなかったことを意味する。さらに、特徴データの粒度に関する問題は、推薦システムにおける一般的な問題であ

るコールドスタート問題とも関係してくる。コールドスタート問題とは、ウェブサービスの開始に推薦の手がかりとなる評価データが非常に少ないため、新たなユーザへの推薦結果の提示や、まだ誰にも評価されていないアイテムを推薦することが困難になるという問題を指す。ソーシャル・デジタルライブラリにおいては、プライバシーの観点からユーザが資料利用履歴に基づく類似読者推薦機能の利用を制限するケースなど、人物推薦機能の対象となるユーザの人数が少ないことが考えられる。その場合に、ユーザの人数に適したユーザ間の共通項を見出すことができない場合には、上述のコールドスタート問題を引き起こす可能性がある。

例えば、新規に構築したソーシャル・デジタルライブラリは、資料利用履歴の蓄積が十分ではなく、利用者も少ないため、コールドスタート問題が発生する可能性がある。このような場合には、特徴データの粒度を粗くすることで類似ユーザを発見しやすくする必要がある。また初期はユーザが少なかったソーシャル・デジタルライブラリも、時間が経過するにつれてユーザの人数が増大していくことが考えられる。その場合には増大したユーザの人数に従って、特徴データの粒度を細かくすることで、類似ユーザを発見する精度を高める工夫が必要である。そのため、ユーザの共通項を発見することに失敗する問題やコールドスタート問題を解決するためには、ソーシャル・デジタルライブラリの状況に従って、特徴データの粒度を柔軟に変更できる必要がある。本論文では、ユーザプロファイリングの際の特徴データを階層的に表現することで粒度を柔軟に変えられるようにする。

また、デジタルライブラリにおける資料利用履歴の特徴として、資料の書誌情報が豊富に提供されていることにある。デジタルライブラリには、書籍や論文などの資料に対応した書誌情報が作成・保存されている。そのため、ソーシャル・デジタルライブラリにおけるユーザの特徴データとして資料利用履歴を活用する際に、ユーザによって直接記録され

た情報だけではなく、記録された資料の書誌情報に着目することでより多くの情報を得ることが可能である。

ソーシャル・デジタルライブラリ上で提供される資料の書誌情報の中で、ユーザの関心を直接表現する要素として、件名標目と分類標目が挙げられる。件名標目とは資料の主題を表現する統制された語彙を掲載する標目である。また、分類標目とは資料を特定の分類法に基づいて分類し、その分類を表現した記号を掲載する標目である。こうした標目は資料の主題を表現する要素であり、その資料を利用したユーザの関心を推測する手がかりとなる。さらに、ユーザの資料利用履歴に紐付いた書誌情報から主題に関する要素を抽出し、集計することによってユーザが利用する資料の主題傾向を把握することができる。

こうした全体的な傾向を把握するための要素としては、件名標目よりも分類標目が適している。資料の分類は、特定の構造を持った分類体系の中に、その資料を位置づける行為である。そのため、集計の際には、分類法が定める分類体系の構造に従って集計することができるため、全体的な傾向を容易に把握することが可能である。件名標目は件名標目表という用語リストの体系に従うが、分類法のような厳密な体系にはなっていない。図 3.1 は、ユーザの関心を資料の分類に基づく階層構造として捉えた場合の図であり、ユーザ A とユーザ B の利用した資料とその分類を階層構造で示している。この場合には、資料の単位では、ユーザ間に共通項が存在しないため、共通の関心を持たないユーザとみなされるが、歴史、日本史といった、資料よりも抽象的な主題を単位とすると、共通項を見出すことができる。このことから、ユーザの関心を推測する手段として、ユーザの資料利用履歴から分類番号を抽出し、抽出した分類番号を集計するアプローチを採用する。

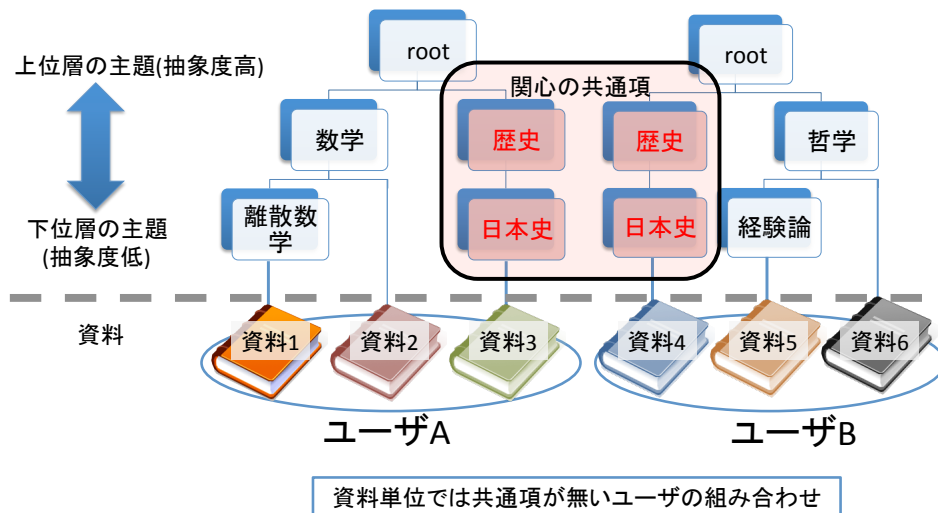


図 3.1 資料の分類に基づく階層構造として捉えたユーザの関心

3.2 NDC とは

ユーザの関心を推測するための特徴データとして分類標目を用いるとき、対象とする分類法の特徴によって集計方法とプロファイルの作成方法は異なってくる。国内の図書館およびデジタルライブラリにおいて、最もよく用いられている分類法に日本十進分類法が挙げられる。日本十進分類法 (NDC; Nippon Decimal Classification, 以下 NDC) とは、日本図書館協会が策定している図書分類法であり、日本の公共図書館の 9 割以上が採用している。NDC は、アメリカ合衆国で策定された分類法である、デューイ十進分類法 (DDC; Dewey Decimal Classification) に基づいて策定されている。そのため、NDC は階層構造を持つことと、特定の分類を 1 桁につき 0~9 までのアラビア数字によって表記することなどの特徴がある。NDC の記号法では、各桁は左から第 1 次区分、第 2 次区分、第 3 次区分..... と呼ばれている。第 3 次区分の桁と第 4 次区分の桁の間にはピリオドを置き、

第4次区分以降は小目として区別する。以下、日本十進分類法に従って表記された分類番号のことをNDC番号と呼ぶ。例えば「日本の近代小説」は913.6と表記する。図3.2は「日本の近代小説」を第1次区分の9類(文学)から階層表示形式で表記した図である。各階層のうち9が「文学」を、91が「文学」のうちの「日本文学」を、913が「日本文学」のうちの「小説. 物語」を、913.6が「日本の小説. 物語」のうちの「日本の近代小説」を指し示している。

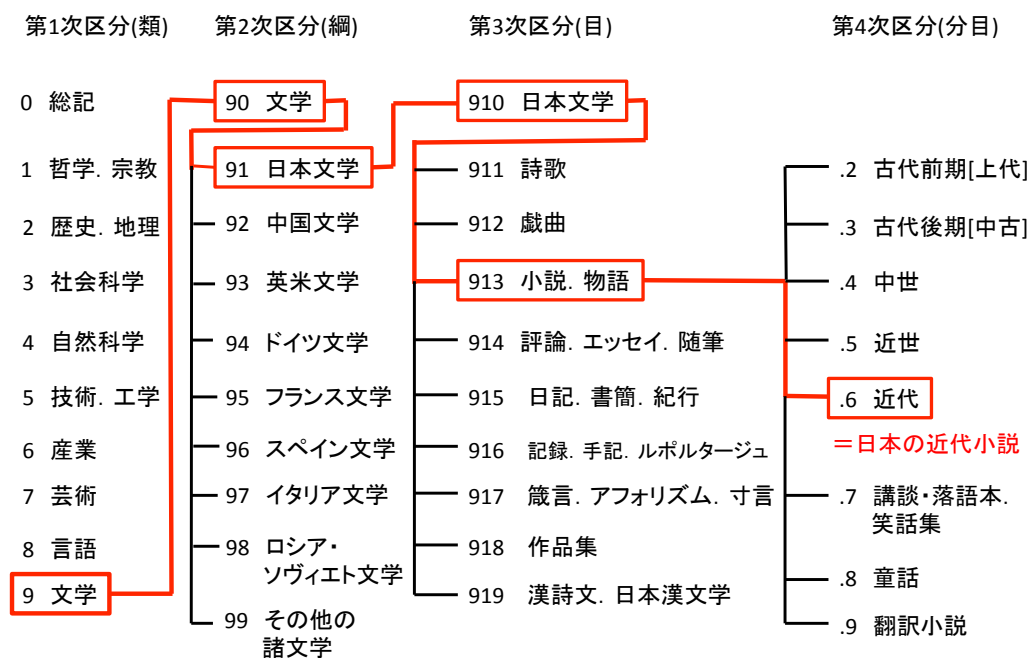


図 3.2 NDC による日本の近代小説の分類 (階層表示形式)

一般に国内で流通している書籍のほとんどには国立国会図書館・公立図書館・MARC提供企業によってNDC番号が付与されている。また、国立国会図書館は、国立国会図書館サーチ (<http://iss.ndl.go.jp/>) という Web サービスにおいて、NDC番号を含む書誌情報を検索できるAPIを提供しており、図書館の環境によらずにNDCを取得することも可能である。そのため、NDC番号は書籍の主題を容易に把握する特徴データとし

て活用できる可能性がある。

さらに、NDC は階層構造をもつ分類法であるため、ある主題に対して、上位に位置する主題や隣接する主題などを推定することにも役立つ。この特性は、利用者の関心の幅や強さを表すために有用である。

ただし、NDC は、本来異なる階層にある主題が同じ階層として扱われていたり、他の主題よりも下位の階層における主題の展開が粗いまたは細かいなど、主題の展開に偏りがある。しかし、こうした NDC の特徴は図書館利用者や現実の資料の傾向を反映したものである。例えば、日本国内では日本文学や英米文学の資料は多数出版されているが、フィリピン文学やアルゼンチン文学などの資料はあまり出版されていない。これはユーザ（読者）の傾向を反映しているともいえる。読者が多く出版点数の多い資料は NDC の上位の階層に位置し、出版点数の少ない資料は NDC の下位に位置している。ユーザ同士の関心を比較するという際に NDC を用いた場合、アルゼンチン文学のような他の分野よりも細分化された分野の資料を読むユーザは資料利用傾向の差異が考慮され、対照的に日本文学のような比較的まとまりのある分野の資料を読むユーザに対しては粒度の粗い主題の一致を高く評価する可能性がある。このように、主題の展開に偏りがある分類法の使用は、むしろ実状にあったプロファイリングを行えることから、人物推薦に有利に働くと考えられる。

3.3 NDC ツリープロファイリング

本論文では、ソーシャル・デジタルライブラリにおけるユーザプロファイリング技術として、NDC ツリープロファイリングを提案する。NDC ツリープロファイリングとは、ユーザの資料利用履歴に含まれる書籍に関する情報から NDC 番号を抽出し、NDC の階層構造に基づいて、ツリー構造で表現したプロファイルを作成する手法である。本手法で

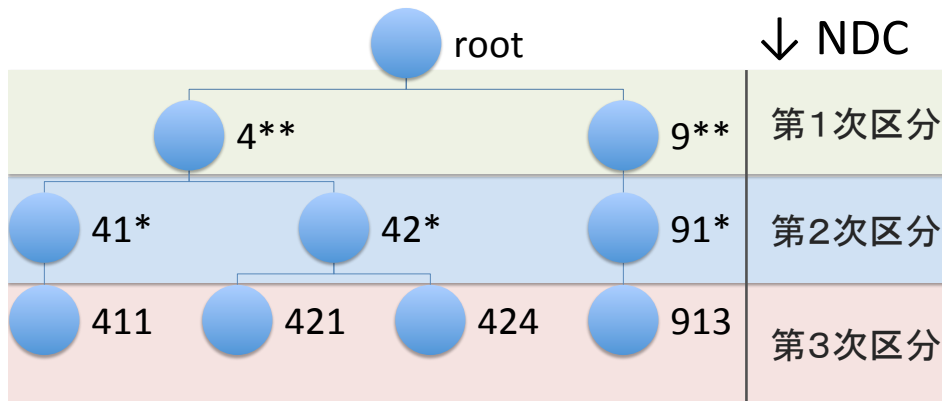


図 3.3 NDC ツリーの例

はこのプロファイルを NDC ツリーと呼ぶ。例えば、あるユーザの読んだ書籍が 5 冊あり、それぞれ「411」「421」「424」「913」「913」と NDC 番号が付与されていたとする。この場合に生成される NDC ツリープロファイルは図 3.3 のようになる。

ここで、3 桁の数字のそれぞれの桁は 3 層に分かれた木構造の区分を示している。例えば、「913」という番号では、1 桁目の 9 は「文学」を表し、2 桁目の「1」は「文学」の中の「日本文学」を、3 桁目の「3」は「日本文学」の中の「小説、物語」を示す。NDC ツリープロファイリングでは、抽出した NDC 番号に対して各階層に対応する桁の番号を表わすノードを作成する。資料利用履歴から抽出された主題は NDC 番号の桁ごとに分けられており、主題分類の粒度を変えてプロファイル間の類似度を測ることができる。本論文ではプロファイルの単純化のため、第 3 次区分を表わす左から 3 桁までをプロファイル作成の対象とし、4 桁目以降は NDC ツリーに反映しないこととする。

なお、NDC 番号からノードを作成する際、既に対応するノードがプロファイルにある場合にはノード作成を行わないこととする。したがって、NDC ツリーのノードはプロファイルの対象ユーザが各ノードを表す主題に関する書籍を最低 1 冊読んでいることを表している。これは特定の資料に対する利用頻度の情報を捨てることを意味している。頻度情報を含めない理由は次のとおりである。NDC ツリープロファイリングは、共通の関

心を持つ利用者を発見し、コミュニティ形成の機会を提供するために提案するユーザプロファイリング手法である。そのため、人物の推薦にあたっては、ユーザの属性にかかわらず、少しでも関心が共通しているユーザを推薦することが望ましい。頻度情報を考慮した場合、多くの資料を読むユーザとあまり本を読まないユーザの類似度が低く算出される可能性がある。このことは、本論文の人物推薦の目的と合致しない結果をもたらす。そのため、頻度情報は NDC ツリーに反映しないこととした。

NDC ツリープロファイリングは、ユーザが読んだ書籍の主題を NDC の階層的な主題分類体系にマッピングする。これによってユーザの読書傾向を表現し、複数の関心によって構成されるユーザの趣味嗜好を表現する。また、NDC ツリープロファイルは階層構造になっているため資料利用履歴の件数が少ないユーザを対象にする場合にも、推薦のための多くの手がかりを得ることに役立つ。そのため、コールドスタート問題を回避できる可能性がある。

NDC ツリープロファイリングの利点は、ユーザの関心を視覚的に表現できる点と、共通の関心を測る際の主題の粒度を調節できる点である。本論文は、ユーザの関心は階層的であるという前提のもと、NDC ツリープロファイリングの提案を通して、階層的なユーザプロファイルを作成することで、ユーザ間の関心の類似性を適切に評価することができるということを主張する。

第 4 章

人物推薦アルゴリズム

4.1 NDC ツリープロファイリングを用いた人物推薦手法

本章では、第 3 章で提案した NDC ツリープロファイリングを利用した人物推薦手法を提案する。本研究で提案する人物推薦では、ユーザプロファイル同士の類似性を測るために、形状類似スコアと共通ラベルスコアの 2 種類の尺度およびコメント率を導入する。そのうえで、これらを組み合わせて、人物推薦を実現するための類似度算出アルゴリズム NDCTREE1 を提案する。さらに、NDCTREE1 の評価実験を行い、その結果から得た知見をもとに改良したアルゴリズム NDCTREE2 を提案する。最後に NDCTREE2 の評価実験を行うことでその有効性を検証する。

NDC ツリープロファイリングを用いた人物推薦の過程は次のとおりである。まず、すべてのユーザの資料利用履歴からユーザごとに NDC ツリーを生成する。次に、NDC ツリーのペアの全ての組み合わせ同士の比較を行う。そして、比較結果を推薦スコアとして算出する。最後に、それぞれのユーザに対して、推薦スコアでランキングしたユーザ上位 5 名を、推薦結果として提示する。

本推薦手法では、NDC ツリーを比較する際の類似尺度として、共通ラベルスコアと形

形状類似スコアを、推薦スコアを調節する補足的な尺度としてコメント率を提案する。

■共通ラベルスコア NDC ツリープロファイリングは、ユーザの関心を NDC という統制された主題分類を用いることで、階層的なプロフィールを作成する手法である。そのため、本研究で提案する人物推薦アルゴリズムには、NDC ツリーの階層毎の共通項を数値化するような尺度を設定することができる。そのような尺度として共通ラベルスコアを提案する。共通ラベルスコアとは、2つの NDC ツリーに対するラベルの一致度からみた資料利用傾向の類似度である。2つのツリーの間で共通するラベルをカウントし、ツリーの階層別に設定された重みをかけ合わせた値の和である。これにより、階層毎に重みを変えて類似度を調整することが可能となる。これは、資料単位で類似度を測る場合や、資料よりも抽象度の高い主題の単位など、ユーザによって類似度を測る粒度を変更するのに適している。なお、共通ラベルスコアはユーザ間の関心の共通項のみを評価し、関心の差異については考慮しない。ここでいう関心の共通項とは、ユーザ同士が共通して関心を抱く主題を指す。一方、関心の差異とは、関心を持つ主題の中で、ユーザ同士が共有しない主題を指す。

たとえ両者の NDC ツリーに大きな差異がみられたとしても、共通項が全てのユーザの組み合わせの中で相対的に多ければ、共通ラベルスコアは高くなり、そのユーザを推薦候補として高く評価する特徴を持つ。関心の差異に対する評価は、後述の形状類似スコアが担う。

■形状類似スコア ユーザ間の関心の差異が大きい場合に、共通項があってもコミュニケーションやコミュニティの形成が妨げられることが考えられる。例えば、図書館情報学という共通の関心を持ったユーザ同士であったとしても、一人は詩について、もう一人は数学についてそれぞれ関心を持ち、コミュニケーションをとりたいユーザであるとする。この際に、関心をもつ主題について相手が興味を示さない場合に、コミュニケーションに

ついて相互に不満を抱き、結果として持続的なコミュニケーションが成り立たないことが考えられる。このような状況を生み出すマッチングを防ぐために、形状類似スコアという尺度を導入する。形状類似スコアは、2つのNDCツリーの形状の差異に着目した資料利用傾向の類似度であり、編集距離と呼ばれる尺度を活用して算出する。編集距離とは、あるツリーを別のツリーと同じ形状とするためにかかる挿入・削除・ラベル編集などの操作のコストを表し、2つのツリーの距離を計算する尺度として一般的に知られている。ツリーの形状の差異が大きいほど編集距離の値は高くなり、2つのツリーが同じ形状を有している場合には編集距離はゼロとなる。形状類似スコアは、ツリーの形状の差異が小さいほど値が高くなるよう変換したスコアである。このようなスコアを導入することで、共通の関心を持ち、関心の差異が少ないユーザを評価するよう調節することが可能になる。

■コメント率 コメント率は、推薦対象のユーザが提供する情報の多寡が、推薦結果を受けるユーザの推薦対象へのコミュニケーションの動機づけに与える影響を示した尺度である。コメント率は、前述の共通ラベルスコアと形状ラベルスコアを補完するための尺度である。形状類似度および共通ラベルスコアは、共にNDCツリーの類似度を測る尺度である。そのため、2つの尺度のみを用いて算出した推薦スコアは、ユーザの組み合わせに対して対称な推薦スコアとなる。これは、ソーシャル・デジタルライブラリにおける人物推薦は、共通の関心を持つユーザを発見することが主な目的であり、資料利用傾向に対するユーザ間の選好は対称性を持つと仮定しているためである。しかし、ユーザがコミュニケーションをとる相手として選好する際には、資料利用傾向以外のユーザに関する情報も影響を与える可能性がある。この資料利用傾向以外の要素としてユーザの参考度が挙げられる。ここでユーザの参考度とは、本研究で独自に想定した要素であり、ユーザがこれまで利用した資料に対して、アノテーションや書評などの付加情報をどれほどソーシャル・デジタルライブラリに付加しているかを表わす。例えば、推薦対象のユーザが、資料利用

履歴に登録されたすべての書籍に対して、書評を投稿していたとする。そのユーザに対しては、より多くの情報や意見を常に提供してくれる有用な情報源として期待が持て、コミュニケーションをとる動機が高まる。一方、まったく情報を付与しないユーザに対しては、すべての書籍に情報を付加しているユーザに比べて提供してくれる情報が相対的に少なくなり、コミュニケーションをとる動機が弱まる。コメント率は、このようなユーザ間でコミュニケーションをとる動機が非対称であることを想定した尺度である。

本論文では上述の3種類の尺度を用いて、推薦スコアの算出式を提案する。

4.2 推薦スコア算出式

4.2.1 推薦スコア算出式の提案

ユーザ i, j の NDC ツリーを T_i, T_j とした場合、共通ラベルスコア SL_{T_i, T_j} は式 4.1 で算出する。

$$SL_{T_i, T_j} = \frac{1}{c_{max}} \sum_{k=1}^3 w_k sl_k(T_i, T_j) \quad (4.1)$$

ここで、

$sl_k(T_i, T_j)$: T_i と T_j の第 k 次区分の階層にある共通ノードの数

w_k : 第 k 次区分の階層の共通ノードに与える重み ($0 \leq w_k \leq 10, k = 1, 2, 3$)

$$c_{max} = \max_{i, j \in U} \sum_{k=1}^3 w_k sl_k(T_i, T_j)$$

U : すべてのユーザの集合

である。

一方、形状類似スコア SF_{T_i, T_j} は、式 4.2 で算出する.

$$SF_{T_i, T_j} = \frac{(d_{max} - d(T_i, T_j))}{d_{max}} \quad (4.2)$$

ここで,

$$d_{max} = \max_{i, j \in U} d(T_i, T_j)$$

$d(T_i, T_j)$: T_i と T_j の編集距離

である. なお、編集距離の計算は Zhang & Shasha が提案しているアルゴリズムを採用した [22].

また、あるユーザ i に対する、コメント率 SC_i は、式 4.3 によって算出される.

$$SC_i = w_{SC} \left(\frac{\text{comment}(i)}{\text{total}(i)} \right) + 1 \quad (4.3)$$

ここで,

$\text{comment}(i)$: ユーザ i の読書記録中にあるコメントの件数

$\text{total}(i)$: ユーザ i の読書記録の件数

w_{SC} : コメント率の重み ($0 \leq w_{SC} \leq 1$)

である.

以上、式 4.2 から 4.3 を組み合わせることによって、ユーザ i をユーザ j に推薦する場合の推薦スコア $\text{score}(i \rightarrow j)$ を、式 4.4 で算出する. なお、 $\text{score}(i \rightarrow j)$ はコメント率を含む非対称なスコアのため、 $\text{score}(j \rightarrow i)$ とは必ずしも同値ではない.

$$\text{score}(i \rightarrow j) = SL_{T_i, T_j} SF_{T_i, T_j} SC_i \quad (4.4)$$

以下、NDC ツリープロファイリングおよび式 4.4 を含めた人物推薦方式を、NDCTREE1 と呼ぶ。

4.2.2 有効性の評価

NDCTREE1 の有効性を検証するために、実際に実験協力者を募り、評価実験を行った。この評価実験を本論文では評価 1 と呼ぶ。

■評価実験の概要 評価実験の内容は次のとおりである。実験期間は 2010 年 12 月 6 日から 13 日の 8 日間である。実験協力者は大学生 37 名である。なお、実験協力に対する報酬などは特に設定していない。評価実験の手順は、実験システム上で実験データを収集する段階と、収集したデータに基づいて、人物推薦のシミュレーションを行う段階の 2 段階に分けた。

実験データを収集する段階では、まず実験協力者に対して、これまでに読んだことのある書籍を、資料利用履歴として実験システム上に登録させた。ただし、この時点では、実験協力者が他の実験協力者の資料利用履歴を閲覧することができないよう機能に制限を加えている。また、実験協力者に対してコメントの投稿方法を教え、書籍についての感想を記録しておきたいときには、書籍を登録する際にコメントも同時に付与することも指示した。

実験協力者の登録作業を完了した後、他の実験協力者の資料利用履歴を閲覧する機能の制限を外し、各実験協力者に全実験協力者の資料利用履歴を閲覧させた。そのうえで、「これから読書をするにあたって参考になる実験協力者を、閲覧した読書記録をもとに 5 名ほど選出してください」と実験協力者に指示した。なお、選出した実験協力者の ID は実験システム上に登録させた。本評価実験では、選出された実験協力者のリストを人物推薦の正解データとする。

次に、人物推薦のシミュレーションを行った。すべての実験協力者の資料利用履歴を入力データとして、すべての実験協力者の組み合わせに対して推薦スコアを算出し、各実験協力者に対して推薦スコア上位5名の実験協力者のリストを、推薦結果として出力した。そのうえで、実験協力者ごとに正解データと推薦結果を比較して評価した。

■実験システム 本評価実験で使用するための実験システムには、Shizuku2.0 という独自に開発したソーシャル・デジタルライブラリを使用した [23]。図 4.1 は、Shizuku2.0 の操作画面である。Shizuku2.0 は、目録を検索し、ユーザがこれまでに読んだ書籍を資料利用履歴^{*1}として登録することができる。なお、資料利用履歴の登録の際には、登録する資料に対してコメントも付与することができるよう機能を実装している。また、他のユーザの資料利用履歴を閲覧する機能も「ウォッチリスト機能」という名称で実装しているが、登録時には閲覧できないよう機能に制限を加えた。

■評価尺度 推薦結果の評価に際しては、正解データに対する精度と再現率、およびその組み合わせの F 値を評価尺度として採用する。精度は、推薦結果として提示された実験協力者のうち、何名が正解データに含まれていたかを示す (式 4.5)。

$$P = \frac{n(\text{rec} \cap \text{correct})}{n(\text{rec})} \quad (4.5)$$

ここで、 rec は推薦結果に含まれる実験協力者の集合を、 correct は正解データに含まれる実験協力者の集合を表わす。また、 $n(x)$ は集合 x の要素数を表わす。なお、本評価では、実験協力者には5名の推薦を行っているため、全て $n(\text{rec}) = 5$ である。

再現率は、正解データに提示された実験協力者のうち、何名が推薦結果に含まれていたかを示す (4.6)。

$$R = \frac{n(\text{rec} \cap \text{correct})}{n(\text{correct})} \quad (4.6)$$

^{*1} Shizuku2.0 内では資料利用履歴を「アクティビティ」と表記している



図 4.1 評価 1: 実験システム 'Shizuku2.0' の操作画面

なお、正解データの数は人によって異なるため、 $n(\text{correct})$ は定数ではない。

F 値は、精度と再現率の釣り合いをあらゆる尺度であり、精度と再現率の調和平均として、式 4.7 で算出する。

$$F = \frac{2 * (P * R)}{P + R} \quad (4.7)$$

■パラメータの最適化 本実験では、NDCTREE1 の 4 つのパラメータの最適化を行った。パラメータは以下の 4 種類を変化させ、精度の平均が最も高いものを求めた。

- 第 1 次区分にあたる共通ラベルの重み w_1 (0~10, 1 刻み)
- 第 2 次区分にあたる共通ラベルの重み w_2 (0~10, 1 刻み)
- 第 3 次区分にあたる共通ラベルの重み w_3 (0~10, 1 刻み)
- コメント率の重み w_{SC} (0.0~1.0 点, 0.1 刻み)

■比較アルゴリズム 比較アルゴリズムとして、推薦対象となる実験協力者をランダムに5人選出するプログラムを用意した。このランダムに選出する比較アルゴリズムをRANDOMと表記する。ユーザ*i*をユーザ*j*に推薦する場合の、RANDOMの推薦スコア算出式を、式4.8として示す。

$$\text{score}(i \rightarrow j) = \text{rand}(0.0, 1.0) \quad (4.8)$$

ここで、 $\text{rand}(x, y)$ は閉空間*x*から*y*までの乱数である。なお、ランダム推薦は実行するたびに推薦結果が変わるため、本研究では100回試行した評価結果の平均値をランダム推薦の評価として採用する。

4.2.3 実験データの集計

実験協力者が登録した書籍の基本統計を表4.1に示す。また、登録された書籍を、NDC第1次区分で分類した場合の内訳を表4.2に示す。実験協力者の中には、272冊(全登録冊数の81.68%)の書籍を登録した者もいれば、まったく登録していない者もいることが分かった。なお、本評価では全く登録していない者も評価対象に加えている。このことから、実験協力者の間には、書籍の登録冊数に大きな差異があることが分かる。

表 4.1 評価 1: 実験協力者が登録した書籍の内訳

全冊数	333 冊
最大値	272 冊
最小値	0 冊
中央値	6 冊
平均値	7.9285 冊
標準偏差	2.9968

正解データの集計結果を表4.3に示す。

表 4.2 評価 1: NDC 第 1 次区分ごとの書籍登録冊数と割合

NDC 第 1 次区分	登録された書籍の割合 (%)
0**(総記)	8 冊 (2.40%)
1**(哲学)	7 冊 (2.10%)
2**(歴史)	1 冊 (0.30%)
3**(社会科学)	12 冊 (3.60%)
4**(自然科学)	6 冊 (1.80%)
5**(技術, 工学)	4 冊 (1.20%)
6**(産業)	1 冊 (0.30%)
7**(芸術, 美術)	9 冊 (2.70%)
8**(言語)	2 冊 (0.60%)
9**(文学)	232 冊 (69.67%)
取得不可	51 冊 (15.32%)
合計	333 冊

表 4.3 評価 1: 正解データの集計

最大値	10
最小値	3
平均	6.1364
標準偏差	1.6867

次に、表 4.2 によれば、NDC 第 1 次区分の分類に基づいた実験協力者全体の資料利用傾向は「文学」を表す 9 類に大きく偏っていることが分かる。なお、取得不可というのは読書記録に記載された書籍の ISBN をもとに、国立国会図書館が提供する書誌の検索システムである国立国会図書館サーチ (<http://iss.ndl.go.jp/>) から NDC 番号の取得を試みたところ、取得に失敗した書籍を指す。このような書籍は、NDC 付与の曖昧性があるため、実験では利用しなかった。

Shizuku2.0 に登録された、ユーザ別のコメント件数を図 4.2 に示した。37 名のユーザうち、10 人が 0 件、6 人が 1 件のみコメントを登録している。

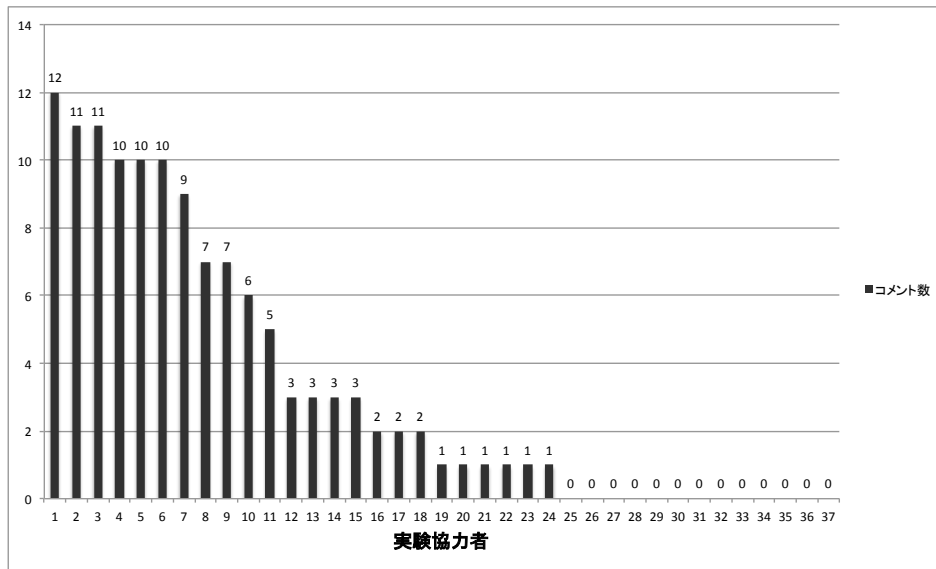


図 4.2 評価 1: Shizuku2.0 に登録されたコメントの件数 (ユーザ別)

表 4.4 評価 1: 各アルゴリズムに対する精度の平均と再現率の平均

推薦手法	精度の平均	再現率の平均
RANDOM	0.1715	0.1722
NDCTREE1	0.2778	0.3556

4.2.4 評価結果

RANDOM, NDCTREE1 の精度の平均と再現率の平均を表 4.4 に示す。ここで NDCTREE1 は、4 つのパラメータを変化させたときの最も良い値を示している。この結果、NDCTREE1 が精度の平均 0.2778, 再現率の平均 0.3556 となり、共に RANDOM の精度の平均, 再現率の平均よりも高くなった。

次に、精度, 再現率が最大の時のパラメータの組み合わせを見たものを表 4.5 に示す。精度の平均と再現率の平均の最大値はそれぞれ 0.2778 と 0.3556 であり、いずれも NDC 第 2 次区分に対する重みであるパラメータ w_2 が、他のパラメータよりも高く設定されており、コメント率のパラメータ w_{SC} はいずれも 0 であった。このことから、精度と再現

表 4.5 評価 1: NDCTREE1 の精度, 再現率が最大の時のパラメータ

w_1	w_2	w_3	w_{SC}	精度の平均	再現率の平均
1	2	1	0	0.2778	0.3556
1	3	0	0	0.2778	0.3556
1	3	1	0	0.2778	0.3556
1	3	2	0	0.2778	0.3556
1	5	0	0	0.2778	0.3556

率の性能にはコメント率は寄与しないことが分かった。

4.3 推薦スコア算出式の改良

4.3.1 改良した算出式

4.2 では, NDC ツリープロファイリングを利用した人物推薦のための推薦スコアの算出式を提案し, その性質および有効性を評価した. その結果, NDCTREE1 は RANDOM よりも高い精度・再現率を示した. しかし, 表 4.5 より, コメント率の精度・再現率への影響はなかった. また, 式 4.4 では, 形状類似スコアに対するパラメータが設定されていないことから, 形状類似スコアと共通ラベルスコアとの関係を調整することが困難である. そこで, NDCTREE1 のスコア算出式 4.4 に対して, コメント率を排除し, 形状類似スコアと共通ラベルスコアの重み付き和として算出するように, 推薦スコア算出式の改良を行った. この人物推薦アルゴリズムを NDCTREE2 と呼ぶ. NDCTREE2 のスコア算出式を式 4.9 に示す. なお, $score(i, j)$ はユーザ i, j に対して対称なスコアであり, $score(j, i)$ と同値になる.

$$score(i, j) = uSF_{T_i, T_j} + (1 - u)SL_{T_i, T_j} \quad (4.9)$$

ここで、 u は SF_{T_i, T_j} に与える 0.0~1.0 の重みであり、形状類似スコアと共通ラベルスコアのバランスを調整するために導入する。

4.3.2 有効性の評価

NDCTREE2 の有効性を検証するために、評価実験を行った。この評価実験を本論文では評価 2 と呼ぶ。

■評価実験の概要 評価実験の内容は次の通りである。実験期間は 2013 年 4 月 11 日から 17 日である。大学生を対象として 24 名の実験協力者を募った。なお、実験協力者に対する報酬は特に設定していない。評価実験の手順は、実験システム上で実験データを収集する段階と、収集したデータに基づいて人物推薦のシミュレーションを行う段階の 2 段階に分けた。それぞれの段階について個別に解説する。

実験データを収集する段階では、実験協力者 24 名に対して、これまでに読んだことのある書籍を実験システム上に登録させ、資料利用履歴を作成した。その際に、実験協力者間での資料利用履歴の冊数に極端な差が出ることをないように、登録冊数を 10 冊から 30 冊の範囲にするよう依頼した。なお、書籍のジャンルは問わなかった。登録作業完了後に、他の全ての実験協力者の資料利用履歴を閲覧させ、それぞれを関心の度合いを基準に、1~5 点の 5 段階で評価するよう指示した。本評価実験では、5 段階評価結果のうち 4 点以上の評価が与えられた実験協力者を正解データとした。なお、正解データの収集作業は、実験協力者 24 名中 19 名の協力があった。正解を得られなかった 5 名の協力者に関しては他の協力者から正解として指定された場合にのみ評価の対象として扱った。

人物推薦のシミュレーションを行う段階では、正解データの収集が行われた 19 名の実験協力者の資料利用履歴を入力データとして、すべての実験協力者の組み合わせに対して推薦スコアを算出した。次に、各実験協力者に対して、推薦スコア上位 5 名を推薦結果

として出力した。そのうえで、実験協力者ごとに正解データと推薦結果を比較して評価した。

推薦結果の評価に際しては、精度・再現率・F 値の人による平均を使用した。

■**実験システム** 本評価実験で使用するための実験システムには、MyBookList という本実験のために独自に開発したソーシャル・デジタルライブラリを使用した。MyBookList は、Shizuku2.0 と同様に、ユーザがこれまでに読んだ書籍を資料利用履歴として登録することができるソーシャル・デジタルライブラリである。ただし、NDCTREE2 のスコア算出式に対応して、Shizuku2.0 にはあった、コメントを付与する機能は実装されていない。図 4.3 は、MyBookList における資料検索・登録画面であり、図 4.4 は、MyBookList における資料利用履歴表示画面である。資料利用履歴表示画面では、ユーザが登録した順に本の書誌情報がリストアップされており、本 1 冊ごとにタイトルと著者、出版社、ISBN、書影が掲載されている。NDC 番号や分類名などの主題に関する情報はユーザには表示されないようになっている。なお、資料利用履歴の登録者本人には、本を資料利用履歴から削除する登録解除ボタンが表示され、他人には非表示となっている。また、ユーザに対する好感度を計測するために、他のユーザを 5 つ星評価方式で評価する機能を実装した。

■**パラメータの最適化** 本実験では NDC ツリープロファイリングによる類似度計算の 4 つのパラメータの最適化を行った。パラメータは以下の 4 種類を変化させ、F 値の平均が最も高いものを求めた。

- 第 1 次区分にあたる共通ラベルの重み w_1 (0~10, 1 刻み)
- 第 2 次区分にあたる共通ラベルの重み w_2 (0~10, 1 刻み)
- 第 3 次区分にあたる共通ラベルの重み w_3 (0~10, 1 刻み)
- 形状類似スコアと共通ラベルスコアのバランスを調整する重み u (0.0~1.0 点, 0.1

刻み)

4.3.3 比較対象の推薦アルゴリズム

比較対象の推薦アルゴリズムとしてランダム推薦 (4.2.2 の式 4.8), 書籍の共通冊数に基づく推薦, NDC 番号の頻度を利用した推薦の 3 種類を実装した。なお, いずれの推薦アルゴリズムも類似度の高い順に作成したランキング上位 5 件を推薦結果として提示する。

書籍の共通冊数に基づく推薦

資料利用履歴中の書籍から, ISBN を抽出したリストをプロフィールとする手法である。ユーザ i とユーザ j の類似度の計算はダイス係数を用い, 以下の式で算出する。

$$score(i, j) = \frac{2 * (v_i \cap v_j)}{|v_i| + |v_j|} \quad (4.10)$$

ここで, v_k はユーザ k の ISBN 番号を要素とする集合である。

書籍の共通冊数に基づく人物推薦は, 既存の読書支援ウェブサービスにも採用されており, 提案手法の既存の手法に対する優位性を検証するために比較手法として採用した, 以後, この手法を BOOK と表記する。

NDC 番号の頻度を利用した推薦

プロフィールを構成するために, 資料利用履歴中の書籍から NDC 番号 (000 から 999 まで, 00*から 99*まで, 0**から 9**まで, 合計 1110 次元) を抽出し, その頻度を特徴量とした, 1110 次元のベクトルをユーザのプロフィールとする手法である。ユーザ i と j の類似度の計算には, 2 つのベクトルのコサイン類似度を用い, 以下の式で算出する。

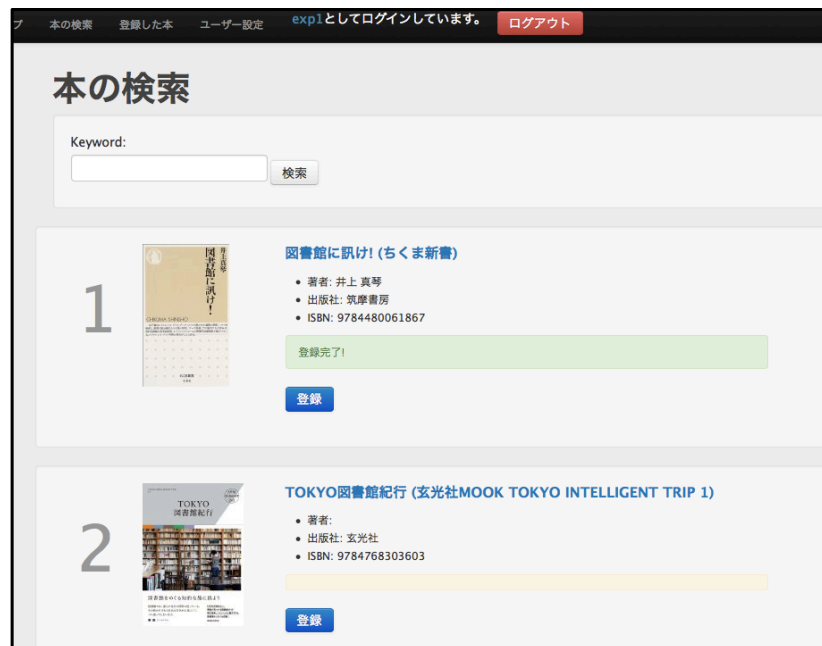


図 4.3 評価 2: MyBookList における資料検索・登録画面



図 4.4 評価 2: MyBookList における資料利用履歴表示画面

$$score(i, j) = \frac{v_i \cdot v_j}{\sqrt{v_i^2} \sqrt{v_j^2}} \quad (4.11)$$

ここで、 v_k はユーザ k の各 NDC 番号の TF-IDF 値を特徴量とする、1110 次元ベクトルである。TF-IDF 値とは、各文書における、ある単語の重要性を算出する重みの一種である。単語の出現頻度を表す TF 値と、その単語が出現する文書の割合を表す IDF 値の 2 種類の尺度を用いて算出する。本アルゴリズムでは、それぞれの NDC 番号に対して、ユーザごとの資料利用履歴中の件数を TF 値、全ユーザの資料利用履歴中にある件数の合計の逆数を IDF 値として計算する。TF-IDF を利用したベクトル空間モデルおよびコサイン類似度による類似度計算手法は、Chen ら [20] が Content-Matching 方式の推薦手法として比較対象に採用していたため本研究でも採用した。以後、この手法を VECTOR と表記する。

4.3.4 実験データの集計

実験協力者が登録した書籍の登録冊数を表 4.6 に示す。実験協力者の登録書籍冊数は、最大で 30 冊 (全登録冊数の 5.70%)、最小で 11 冊である。このことから、表 4.1 で示した改善前の評価と比較して、書籍の登録冊数の差異は小さい。

表 4.6 評価 2: 実験協力者が登録した書籍の内訳

全冊数	526 冊
最大値	30 冊
最小値	11 冊
平均	21.9166 冊
中央値	24 冊
標準偏差	7.0587

また、登録された書籍を NDC 第 1 区分で分類した場合の内訳を表 4.7 に示す。なお、

取得不可というのは、資料利用履歴に記載された書籍の ISBN をもとに、国立国会図書館サーチから NDC 番号の取得を試みたところ、取得に失敗した書籍を指す。この書籍は、NDC 付与の曖昧性があるため、実験では利用しなかった。

表 4.7 評価 2: NDC 第 1 次区分ごとの書籍登録冊数と割合

NDC 第 1 次区分	登録された書籍の割合 (%)	
0** (総記)	29 冊	(5.51%)
1** (哲学)	5 冊	(0.95%)
2** (歴史)	8 冊	(1.52%)
3** (社会科学)	21 冊	(4.00%)
4** (自然科学)	22 冊	(4.18%)
5** (技術, 工学)	11 冊	(2.09%)
6** (産業)	4 冊	(0.76%)
7** (芸術, 美術)	62 冊	(11.79%)
8** (言語)	2 冊	(0.38%)
9** (文学)	354 冊	(67.30%)
取得不可	8 冊	(1.52%)
合計	526 冊	

正解データの集計結果を表 4.8 に示す。

表 4.8 評価 2: 正解データの集計

最大値	13
最小値	1
平均	5.3684
標準偏差	3.6159

4.3.5 評価結果

NDCTREE2 と比較アルゴリズムそれぞれに対する、精度の平均、再現率の平均、F 値の平均を表 4.9 に示す。これにより、RANDOM, BOOK, VECTOR と比べて、

表 4.9 評価 2: 各アルゴリズムに対する精度の平均と再現率の平均

アルゴリズム	精度の平均	再現率の平均	F 値の平均
RANDOM	0.2366	0.2227	0.2294
BOOK	0.3158	0.2839	0.2990
VECTOR	0.2842	0.2539	0.2681
NDCTREE2	0.3473	0.3382	0.3427

表 4.10 評価 2: NDCTREE2 の評価結果上位 11 件のパラメータ

w_1	w_2	w_3	u	精度の平均	再現率の平均	F 値の平均
1	4	0	0.1	0.3473	0.3382	0.3427
1	5	0	0.1	0.3473	0.3382	0.3427
1	6	0	0.1	0.3473	0.3382	0.3427
2	8	0	0.1	0.3473	0.3382	0.3427
2	9	0	0.1	0.3473	0.3382	0.3427
2	10	0	0.1	0.3473	0.3382	0.3427
1	0	0	0	0.3157	0.3703	0.3409
2	0	0	0	0.3157	0.3703	0.3409
3	0	0	0	0.3157	0.3703	0.3409
4	0	0	0	0.3157	0.3703	0.3409
5	0	0	0	0.3157	0.3703	0.3409

NDCTREE2 の精度の平均, 再現率の平均, F 値の平均が高いことが分かった。

次に, パラメータの傾向を見るために, F 値の平均の上位 11 件について, すべてのパラメータの組み合わせと再現率・精度・F 値それぞれの平均を, 表 4.10 に示す。精度の平均と再現率の平均の最大値は, それぞれ 0.3473 と 0.3382 であり, いずれもパラメータ w_2 の値を w_1, w_3 と比較して高く設定した場合であった。

図 4.5 は, 全実験協力者の F 値の平均を最大化するパラメータ ($w_1 = 1, w_2 = 4, w_3 = 0, u = 0.1$) のときの, ユーザ別の F 値を棒グラフで示した図である。F 値の最大値は 0.5714 であり, 最小値は 0.0 である。また, F 値が 0.0 となったユーザは 4 名いた。

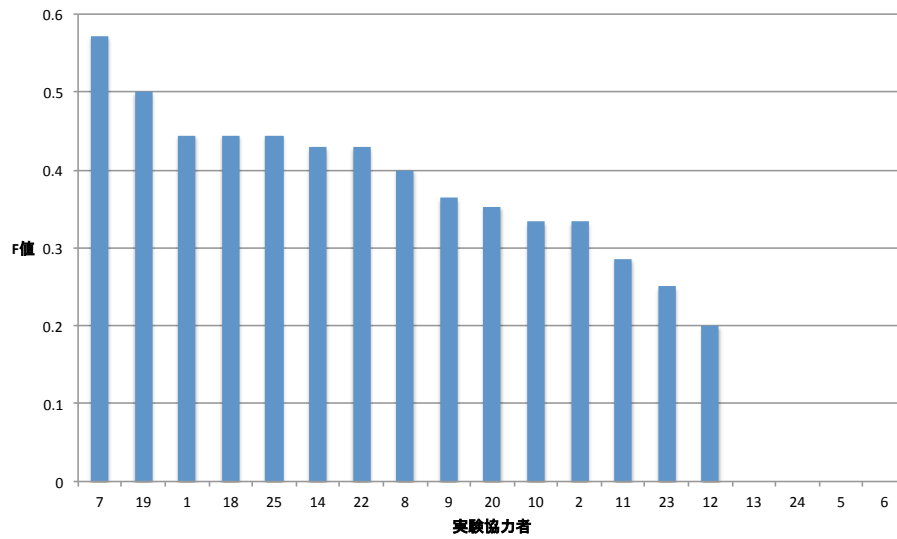


図 4.5 評価 2: NDCTREE2 のユーザ別の F 値

次に、NDCTREE2 の 4 つのパラメータの性質を観察するために、各パラメータごとに、F 値の平均値の推移を図 4.6 に示す。図中のグラフは、全実験協力者の F 値の平均を最大化するパラメータ ($w_1 = 1, w_2 = 4, w_3 = 0, u = 0.1$) のうち、3 つを固定し、残り 1 つのパラメータの値を変化させた時の、F 値の推移を示している。 w_1 に対する F 値の平均値は、 $w_1 = 1$ のとき最大値 0.3042 を、 $w_1 \geq 6$ のとき最小値 0.2485 を示している。 w_2 に対する F 値の平均値は、 $4 \leq w_2 \leq 6$ のとき最大値 0.3042 を、 $w_2 = 1$ のとき最小値 0.2515 を示している。 w_3 に対する F 値の平均値は、 $w_3 = 0$ のとき最大値 0.3042 を、 $w_3 = 7, 8$ のとき最小値 0.2172 を示している。 u に対する F 値の平均値は、 $u = 0.1$ のとき最大値 0.3042 を、 $u = 0.5$ のとき最小値 0.2297 を示している。

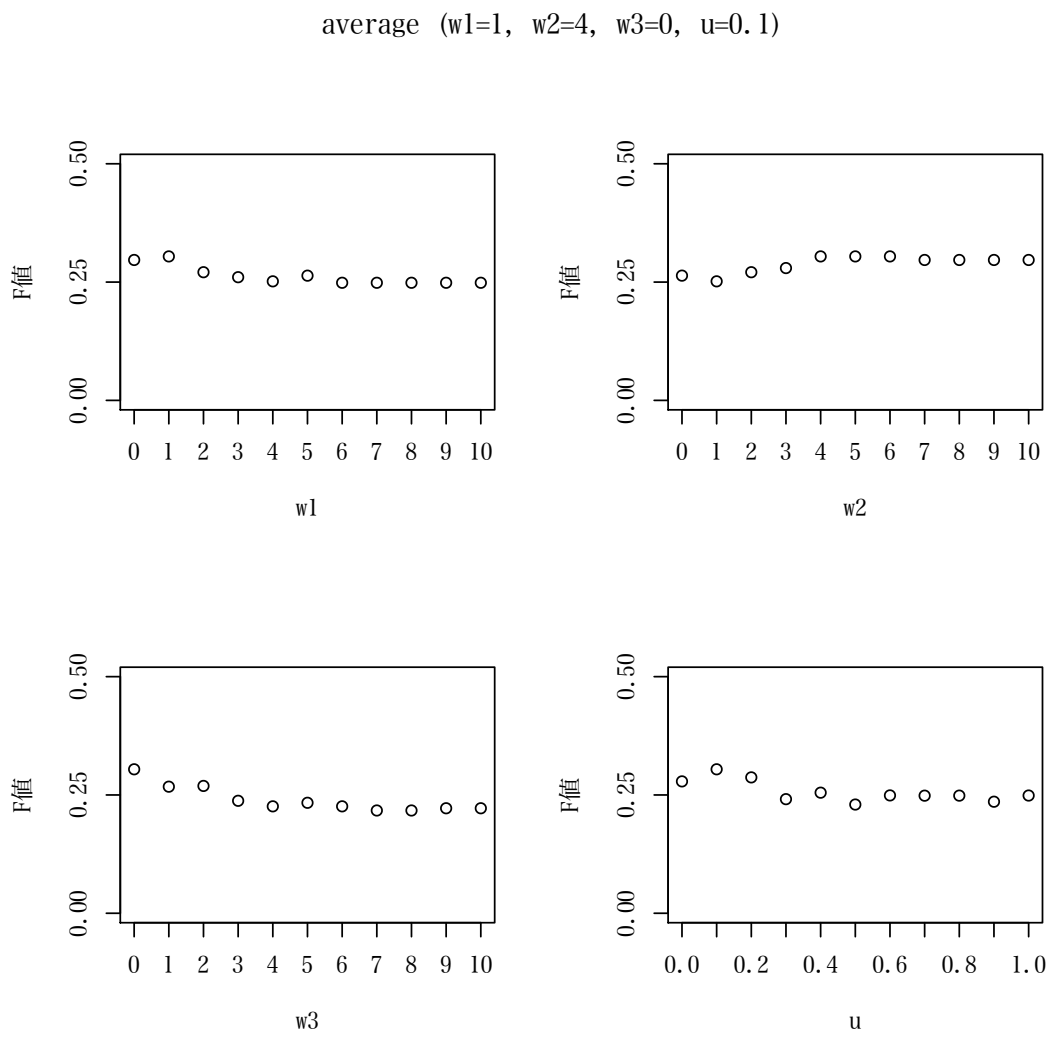


図 4.6 評価 2: パラメータに対する F 値の平均値の推移

第 5 章

考察

5.1 NDC ツリープロファイリングの有効性

4.3.2 における評価結果から、NDCTREE2 は比較アルゴリズムである RANDOM よりも再現率・精度が高いことが示された。このことから、本手法は人物推薦として意味があることが分かった。さらに、主題の階層性を考慮しない推薦方式である VECTOR と BOOK よりも精度・再現率が高いことが示され、パラメータの最適化において w_2 を重視することで結果が最適化されることが示された。このことから、階層性を考慮に入れた主題の共通性が人物推薦に良い結果をもたらすという本研究の仮説が検証された。

5.2 主題の粒度の影響

NDC 番号を特徴量として用いる課題として、小説に対する分類でジャンルが考慮されないことが挙げられる。NDC では小説は 9 類 (文学) に分類され、国別に 913, 923, 933, … と NDC 番号が割り振られる。小説には、一般には恋愛小説やホラー小説などといったジャンルの区分が存在するが、NDC はその区分を採用していない。そのため、NDC

番号に基づいて資料利用履歴から資料利用傾向を分析した場合、恋愛小説のみが好きな読者とホラー小説のみが好きな読者を、プロフィール上では区別できないという問題が起こる。資料利用傾向の類似性を測るという意味においては、これは人物推薦の精度を低下させる要因になる可能性がある。

その一方で、ジャンルにこだわらず小説を読むかどうか、自分にとって参考とする人物の基準と考える人にとっては、あまり問題にならないことも考えられる。本研究で収集した資料利用履歴中の書籍のうち、60%超が小説であり、既存の読書系 SNS においても、小説が資料利用履歴の多くを占める傾向があるため、精度などに大きく影響を与える。

以上のことから、分類の粒度が粗い小説を除外した場合の評価を行い、小説が本手法に与える影響を調査した。具体的には、4.3.2 で収集した資料利用履歴から、「近代以降の日本の小説 (NDC:913.6)」に分類される資料を除外したうえで、NDCTREE2 の精度・再現率・F 値を求めた。また、パラメータに与える影響も考察するために、F 値のパラメータによる変化も検証した。ここでは、元の資料利用履歴に対して NDCTREE2 を適用した場合 ALL と呼び、除外処理後の資料利用履歴に対して NDCTREE2 を適用した場合 EXCLUDE9136 と呼ぶ。

ALL と EXCLUDE9136 の精度、再現率、F 値の平均を表 5.1 に示す。精度の平均は ALL で 0.3473、EXCLUDE9136 で 0.3052 であり、除外後の精度が低下していることを示している。一方、再現率の平均は ALL で 0.3382、EXCLUDE9136 で 0.3618 であり、除外後の再現率が向上している。F 値の平均は、ALL で 0.3427、EXCLUDE9136 で 0.3310 であり、除外後のほうが F 値は下がっている。

表 5.1 小説を除外することによる精度・再現率・F 値の差異

資料利用履歴	精度の平均	再現率の平均	F 値
ALL	0.3473	0.3382	0.3427
EXCLUDE9136	0.3052	0.3618	0.3310

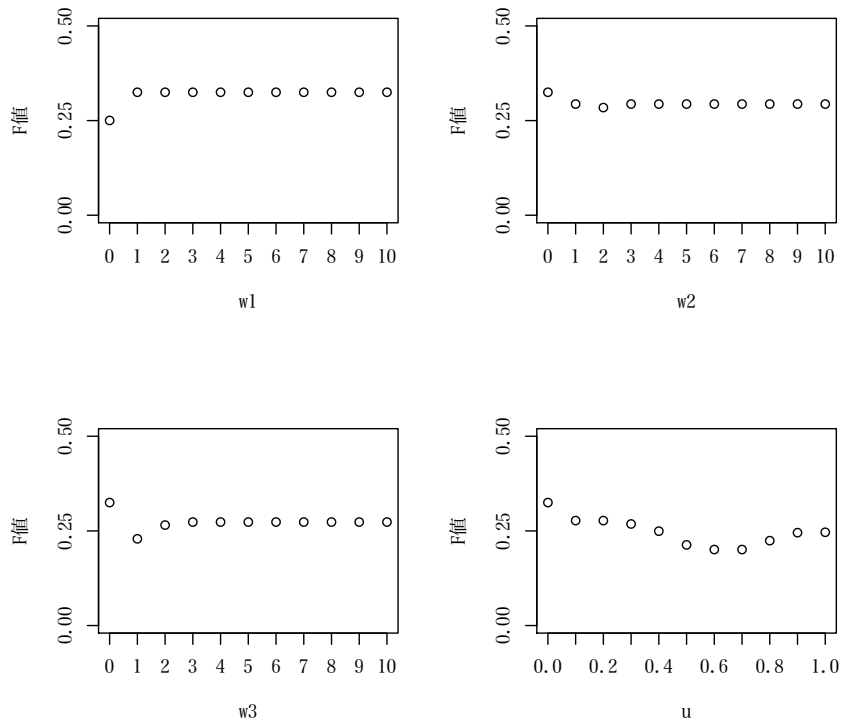
EXCLUDE9136 ($w_1=1, w_2=0, w_3=0, u=0$)

図 5.1 913.6 を除外した場合のパラメータに対する F 値の推移

ALL に対する結果と、各パラメータの性質について比較するために、EXCLUDE9136 に対する結果の、各パラメータごとの F 値の最大値の推移を図 5.1 に示す。 w_1 については、 $1 \leq w_1$ のとき F 値が最大を、 $w_1 = 0$ のとき F 値が最小を示している。 w_2 については、 $w_2 = 0$ のとき F 値が最大を、 $w_2 = 2$ のとき F 値が最小を示している。 w_3 については、 $w_3 = 0$ のとき F 値が最大を、 $w_3 = 1$ のとき F 値が最小を示している。 u については、 $u = 0$ のとき F 値が最大を、 $u = 0.6, 0.7$ のとき F 値が最小を示している。

小説を抜いたデータによる評価結果において、最適化の結果として小説を含めた場合と同等の性能であった。これは、重要視する主題の粒度を変更することで達成できたことである。このことから、本手法では利用する資料利用履歴の傾向に応じて主題の粒度を調節

することができることが分かった。ただし、形状類似スコアの重みは 0.0 で最適化されることから、ユーザ間の資料利用傾向の差異を考慮するアプローチが有効に働いていないことを示している。

5.3 実験協力者ごとの差異

4.3.5 では、NDCTREE2 のパラメータに対する、全実験協力者の F 値を平均した値の推移を示し、F 値が最大となるパラメータを求めた。しかし、全ての実験協力者に対して同一のパラメータが F 値を最大化するとは限らない。そこで、各実験協力者の F 値がどのようにになっているかを検証した。

4.3.2 の評価実験における、NDCTREE2 の精度、再現率、F 値、F 値の最大値を実験協力者ごとに表 5.2 に示す。ここで、正解数とは、実験協力者が関心があったとした他の実験協力者の人数である。また NDC1 は、実験協力者の NDC ツリーにおける、NDC 第 1 次区分に該当するノード数である。NDCTREE2 の再現率・精度・F 値は、すべての実験協力者に対する、F 値の平均が最大になるようなパラメータ $w_1 = 1, w_2 = 4, w_3 = 0, u = 0.1$ を指定した時の数値を示している。一方、F 値の最大値は実験協力者ごとに F 値が最大となるようにパラメータを最適化して得た値である。なお、再現率・精度がともに 0.0 である場合の F 値は、平均の算出のために便宜上 0.0 としている。

ここで、表 5.2 から、全実験協力者の F 値を最大化するパラメータが必ずしも実験協力者の F 値を最大化しないことが分かる。例えば、実験協力者 14 番の場合、全体を最適化したときの F 値は 0.4286 だが、実験協力者 14 番を最適化した場合の F 値は 0.5714 である。

パラメータの変化に対する、実験協力者別の F 値の最大値が、全実験協力者に対する F 値の平均の最大値と異なる例として、実験協力者 14 番を見る。推薦結果の F 値は、

表 5.2 実験協力者ごとの NDCTREE2 の精度・再現率・F 値

実験協力者	正解数	NDC1	NDCTREE2			
			再現率	精度	F 値	F 値の最大値
1	4	3	0.5000	0.4000	0.4444	0.6666
2	7	5	0.2857	0.4000	0.3333	0.5000
5	2	3	0.0000	0.0000	0.0	0.2857
6	2	2	0.0000	0.0000	0.0	0.2857
7	2	2	1.0000	0.4000	0.5714	0.5714
8	9	4	0.4000	0.4000	0.4000	0.6000
9	6	6	0.3333	0.4000	0.3636	0.5454
10	1	4	1.0000	0.2000	0.3333	0.3333
11	9	6	0.2222	0.4000	0.2857	0.5714
12	5	2	0.2000	0.2000	0.2000	0.6000
13	1	3	0.0000	0.0000	0.0	0.3333
14	9	1	0.3333	0.6000	0.4286	0.5714
18	4	2	0.5000	0.4000	0.4444	0.4444
19	7	3	0.4286	0.6000	0.5000	0.5000
20	12	5	0.2500	0.6000	0.3529	0.5882
22	9	3	0.3333	0.6000	0.4286	0.4285
23	3	9	0.3333	0.2000	0.2500	0.5000
24	1	3	0.0000	0.0000	0.0	0.3333
25	13	6	0.3077	0.8000	0.4444	0.4444
平均			0.3383	0.3474	0.3043	0.4791

パラメータは $w_1 = 0, w_2 = 0, w_3 = 0, u > 0$ の時または、 $u \geq 0.4$ である場合に最大値 0.5714 を示している。パラメータ u の値が 0.4 以上のとき、F 値が最大値 0.5714 を示しており、全実験協力者の F 値の平均値とパラメータの関係性とは異なる傾向を示している。次に、実験協力者 14 番のパラメータの推移を図 5.2 に示す。図中のグラフは、実験協力者 14 番に対して、全実験協力者の F 値の平均を最大化するパラメータ ($w_1 = 1, w_2 = 4, w_3 = 0, u = 0.1$) を付与した上で、それぞれのパラメータの値を変化さ

せた時の F 値の推移を示している。 u の値が高いほど、形状類似スコアを重視した推薦結果となることから、実験協力者 14 番の他の実験協力者に対する選好を、形状類似スコアの尺度がよく説明できていることがわかる。

実験協力者14番 ($w_1=1, w_2=4, w_3=0, u=0.1$)

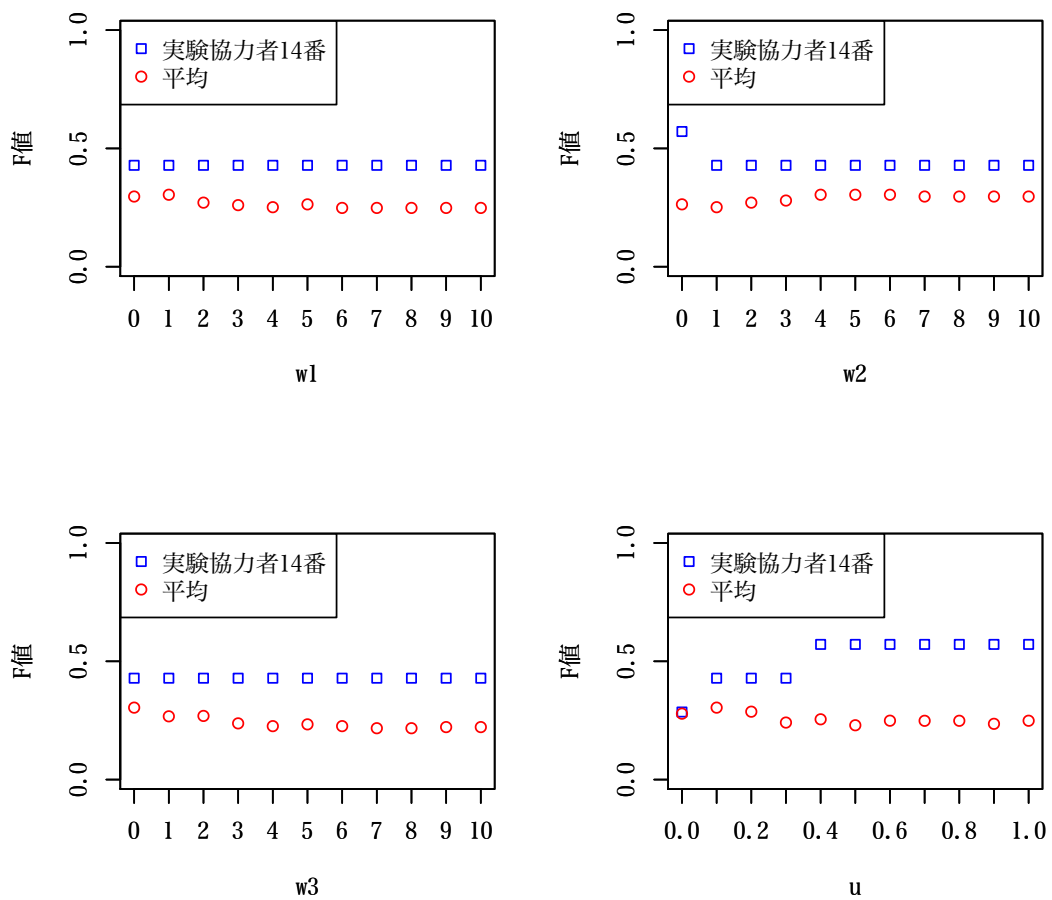


図 5.2 実験協力者 14 番のパラメータに対する F 値の推移

このことを具体的に示すために、実際のツリー構造を確認した。図 5.3 は、実験協力者 14 番の F 値が最大になるように推薦結果を最適化した際に、実験協力者 14 番と共に実験協力者 14 番に推薦される実験協力者 7 番、実験協力者 12 番の NDC ツリーである。な

お、実験協力者 7 番および実験協力者 12 番は実験協力者 14 番が選好した実験協力者でもある。実験協力者 14 番の資料利用履歴は、すべて文学に関する書籍であり、NDC ツリーは NDC 第 1 次区分に該当するノード数が 1 であり、文学の分類である 9 類のみで構成されている。このことから、実験協力者 14 番の関心の幅は他と比較して狭いことが推測される。実験協力者 7 番、実験協力者 12 番の NDC ツリーもまた同様に、NDC 第 1 次区分のノード数が少なく、9 類を重視した形状となっている。実験協力者 14 番は、他の実験協力者と共有できる関心が少ないため、共通ラベルスコアによって類似の実験協力者を発見することが難しい。一方で、形状類似スコアは、ユーザ同士の関心の差異に着目する類似尺度であり、関心の差異が大きくなならないユーザの推薦スコアを高める特徴がある。実験協力者 14 番に対して、パラメータ u を重視した推薦の F 値が比較的高くなったのは、実験協力者 14 番と実験協力者 14 番が選好するユーザの関心の幅が共に狭く、関心の差異が大きくなならないような組み合わせであったことが影響していると推測される。

次に、実験協力者 11 番と NDCTREE2 によって推薦された実験協力者のうち、ユーザが選好した実験協力者の NDC ツリーを図 5.4 に示し、さらに実験協力者 11 番のパラメータの推移を図 5.5 に示す。実験協力者 11 番の NDC ツリーでは、NDC 第 1 次区分に該当するノード数が比較的多い。推薦された実験協力者の NDC ツリーでも同様に、NDC 第 1 次区分に該当するノード数が多い。また、図 5.5 から、パラメータ u に対する F 値の推移は $0.0 \leq u \leq 0.6$ で最大値となり、 $u \geq 0.7$ のとき F 値の値が減少している。すなわち、今回の実験では、幅広い関心を持つ実験協力者である実験協力者 11 番に対して、形状類似スコアはあまり効果が無いことが示されている。

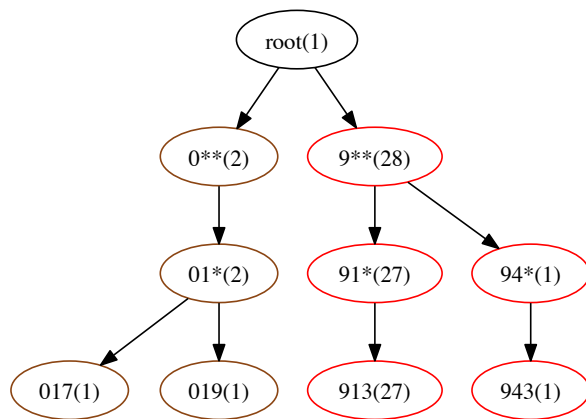
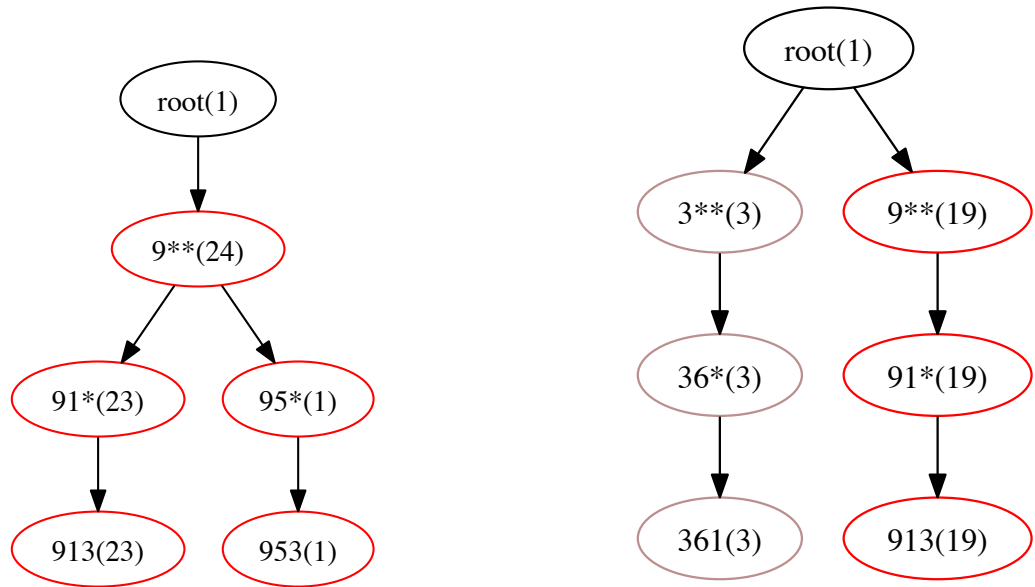
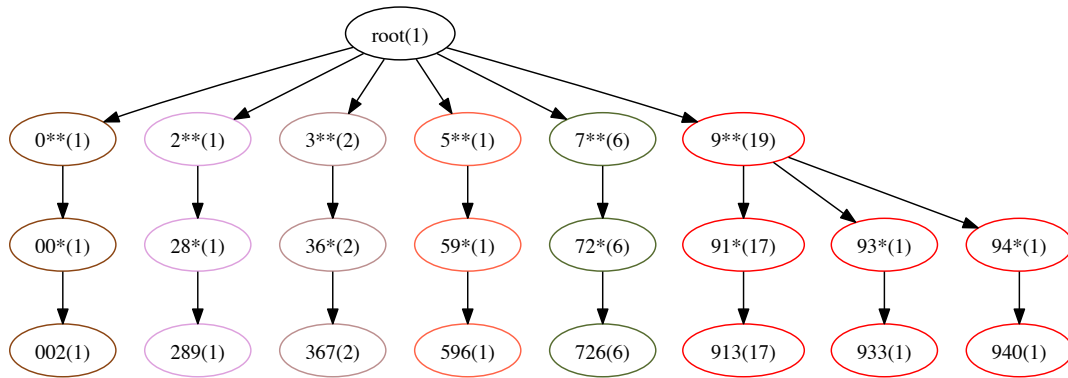
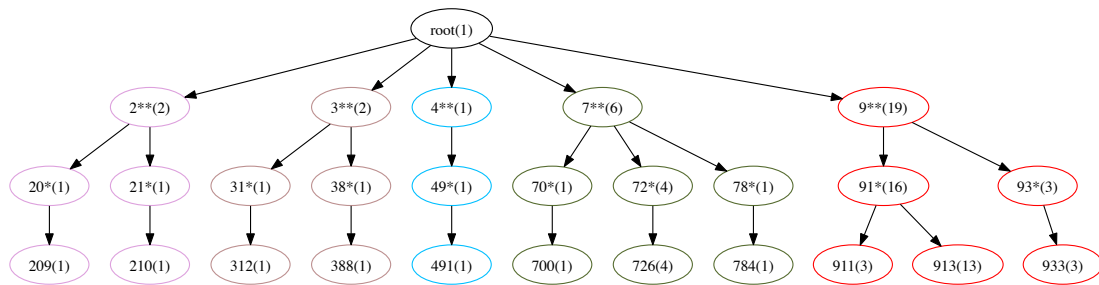


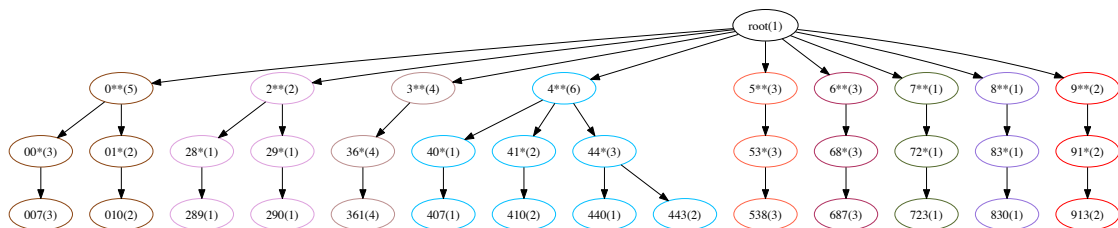
図 5.3 実験協力者 14 番と推薦された実験協力者の NDC ツリー



(a) 実験協力者 11 番の NDC ツリー



(b) 実験協力者 2 番の NDC ツリー



(c) 実験協力者 23 番の NDC ツリー

図 5.4 実験協力者 11 番と推薦された実験協力者の NDC ツリー

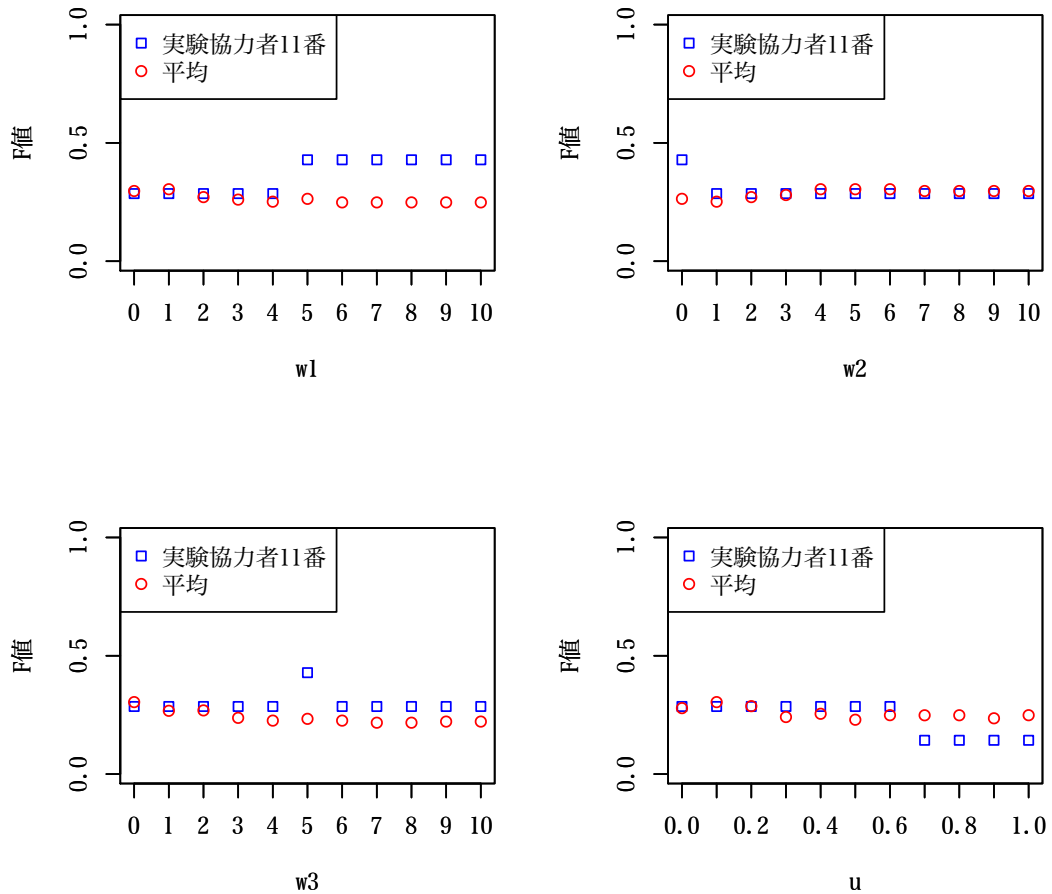
実験協力者11番 ($w_1=1, w_2=4, w_3=0, u=0.1$)

図 5.5 実験協力者 11 番における、パラメータに対する F 値の推移

5.4 NDCTREE2 と BOOK の比較

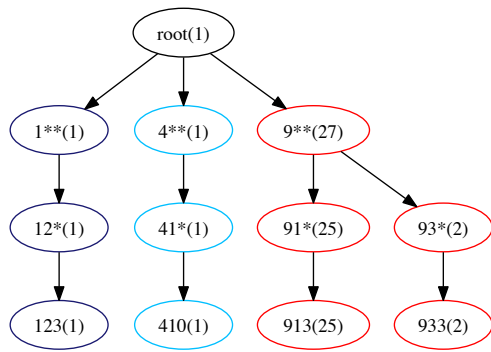
実験協力者ごとのアルゴリズムの比較を表 5.3 に示す。NDCTREE2 の性能が BOOK よりも高い性能を示したのは、全実験協力者のうち 7 名であった。対照的に、BOOK の性能が NDCTREE2 よりも高い性能を示したのは、全実験協力者のうち 5 名であった。このうち、NDCTREE2 のほうが高い性能を示した事例として実験協力者 19 番を、BOOK

表 5.3 実験協力者ごとのアルゴリズムの比較

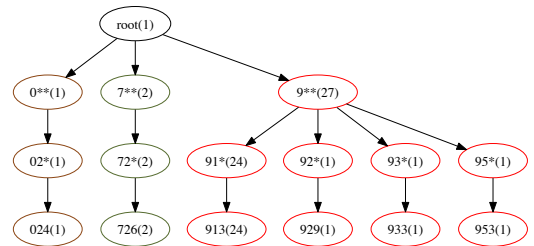
実験協力者	NDCTREE2			VECTOR			BOOK		
	再現率	精度	F 値	再現率	精度	F 値	再現率	精度	F 値
1	0.5000	0.4000	0.4444	0.7500	0.6000	0.6667	0.7500	0.6000	0.6667
2	0.2857	0.4000	0.3333	0.4286	0.6000	0.5000	0.4286	0.6000	0.5000
5	0.0000	0.0000	0.0	0.5000	0.2000	0.2857	0.5000	0.2000	0.2857
6	0.0000	0.0000	0.0	0.0000	0.0000	0.0	0.0000	0.0000	0.0
7	1.0000	0.4000	0.5714	0.5000	0.2000	0.2857	0.0000	0.0000	0.0
8	0.4000	0.4000	0.4000	0.4000	0.4000	0.4000	0.2000	0.2000	0.2000
9	0.3333	0.4000	0.3636	0.3333	0.4000	0.3636	0.3333	0.4000	0.3636
10	1.0000	0.2000	0.3333	0.0000	0.0000	0.0	0.0000	0.0000	0.0
11	0.2222	0.4000	0.2857	0.2222	0.4000	0.2857	0.3333	0.6000	0.4286
12	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000
13	0.0000	0.0000	0.0	0.0000	0.0000	0.0	1.0000	0.2000	0.3333
14	0.3333	0.6000	0.4286	0.2222	0.4000	0.2857	0.3333	0.6000	0.4286
18	0.5000	0.4000	0.4444	0.5000	0.4000	0.4444	0.2500	0.2000	0.2222
19	0.4286	0.6000	0.5000	0.1429	0.2000	0.1667	0.2857	0.4000	0.3333
20	0.2500	0.6000	0.3529	0.2500	0.6000	0.3529	0.2500	0.6000	0.3529
22	0.3333	0.6000	0.4286	0.2222	0.4000	0.2857	0.2222	0.4000	0.2857
23	0.3333	0.2000	0.2500	0.0000	0.0000	0.0	0.0000	0.0000	0.0
24	0.0000	0.0000	0.0	0.0000	0.0000	0.0	0.0000	0.0000	0.0
25	0.3077	0.8000	0.4444	0.1538	0.4000	0.2222	0.3077	0.8000	0.4444
平均	0.3383	0.3474	0.3043	0.2540	0.2842	0.2497	0.2839	0.3158	0.2655

のほうが高い性能を示した事例として実験協力者 11 番を個別に分析し、NDCTREE2 と BOOK の性能について比較する。

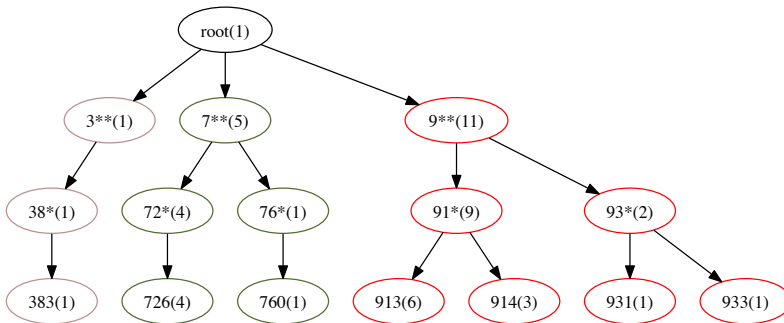
実験協力者 19 番と NDCTREE2 によって推薦された実験協力者の NDC ツリーを図 5.6 に示す。なお、推薦された実験協力者は、すべて実験協力者 19 番に選好されている。このうち、実験協力者 5 番は BOOK も推薦しており、実験協力者 25 番と実験協力者 13 番は NDCTREE2 のみが推薦している。実験協力者 19 番は 4 名の実験協力者とそれぞれ 1 冊ずつ書籍を共有しており、このうち実験協力者 2 番と実験協力者 13 番の 2 名が実験協力者 19 番が選好する実験協力者である。また、実験協力者 19 番が選好した、実験協力者 25 番および実験協力者 13 番の共通書籍冊数は 0 冊である。ここで、実験協力者 25



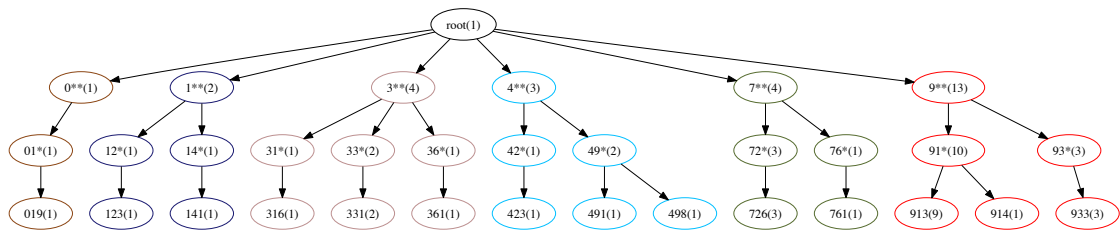
(a) 実験協力者 19 番の NDC ツリー



(b) 実験協力者 5 番の NDC ツリー



(c) 実験協力者 13 番の NDC ツリー



(d) 実験協力者 25 番の NDC ツリー

図 5.6 実験協力者 19 番と、推薦された実験協力者の NDC ツリー

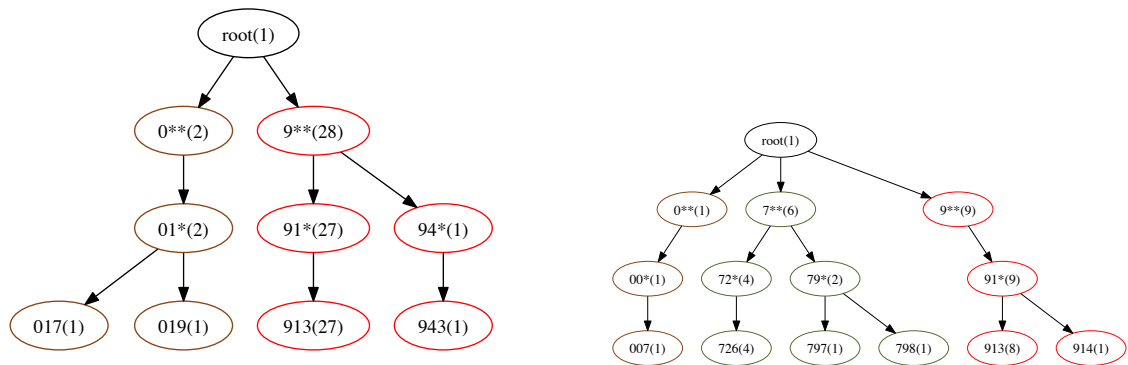
番は共通書籍冊数は0冊であるが、文学以外の分野のノードが多数共通しており、BOOKでは捉えられない共通の関心を捉えることに成功している。実験協力者19番ではNDCツリーにおける、NDC第1次区分のノードの数が類似している実験協力者が、推薦される傾向にある。このことから、共通ラベルスコアは、実験協力者19番のような、小説以外の分野にも関心がある実験協力者に対して有効である。

一方、実験協力者11番はBOOKのF値がNDCTREE2と比較して高く、提案手法が効果的ではないことを示している。実験協力者11番は2名の実験協力者と2冊ずつ、3名の実験協力者と1冊ずつ書籍を共有している。このうち3名は実験協力者11番が選好する実験協力者であり、共通冊数が2冊の実験協力者はすべてこの中に入っている。2冊以上の書籍を共有する実験協力者の組み合わせは少なく、BOOKがNDCTREE2よりも高いF値を示しているのは、実験協力者11番の選好が、実験協力者との共通書籍冊数に影響を受けていると推測される。実験協力者11番が選好し、BOOKが推薦したが、NDCTREE2は推薦しなかった実験協力者のNDCツリーを図5.7に示す。図5.4のNDCTREE2が推薦した、幅の広い形状を備えたNDCツリーを持つ実験協力者とは異なり、幅の狭い形状である。これらの協力者は、BOOKが推薦していることから、特定の本による共通項によって実験協力者11番に選好された例であると考えられる。

5.5 NDC ツリープロファイリングの課題

5.5.1 実験条件の課題

今回の実験は、NDCTREE2に対する評価実験の実験協力者数は24名であり、実験協力者の属性は筑波大学図書館情報メディア研究科の学生のみであるという限定的な条件での結果であったが、形状類似スコアの有効性はユーザの関心の幅の広さに左右されること



(a) 実験協力者 12 番の NDC ツリー

(b) 実験協力者 1 番の NDC ツリー

図 5.7 実験協力者 11 番が選好し、BOOK のみが推薦した NDC ツリー

や、BOOK と比較して提案手法が高い F 値を示すことを明らかにした。更なる検証のためには、実験協力者数が十分大きく、様々な資料利用傾向を持つ実験協力者の集合に対して評価し、詳細なアンケートやインタビューを付随して行うなど、実験条件を改善する必要がある。

5.5.2 類似尺度の妥当性

形状類似スコアの重み u は、NDCTREE2 の評価実験では 0.1、日本の近代小説を除いた追加評価では 0.0 で F 値が最大値を示した。このことから、全体的な傾向として形状類似スコアが人物推薦の性能に与える影響はわずかであるといえる。4.1 で示したように、形状類似スコアはユーザ間の NDC ツリーの差異に着目する類似尺度である。対して、共通ラベルスコアは NDC ツリーの共通項のみに着目する類似尺度である。形状類似スコアの影響がわずかであるという評価結果は、本評価実験の条件下ではユーザの人物に対する選好は関心の差異に大きく左右されなかったことを意味している。ただし、5.3 で述べた

ように、形状類似スコアは関心の広がり狭い実験協力者には有効な結果を示している。このことから、形状類似スコアが影響を与える場合と与えない場合を比較した上で、算出式の更なる改善が必要である。例えば、本論文では、NDC ツリーの各ノードに対して資料の冊数は記録されていない。資料の冊数を NDC 番号が指す主題に対する関心の重みとして利用することで、更なる性能の改善ができることが考えられる。

形状類似スコアに対する以上の知見は、ソーシャル・デジタルライブラリ上のユーザの関係やコミュニティの形態の特徴を示唆している。第 1 章で述べたように、ソーシャル・デジタルライブラリは、資料をバウンダリオブジェクトとして提供することで、ユーザコミュニティを細分化または明示化することを意図するデジタルライブラリである。もし本研究で示された、ユーザ間の関心の差異に対する影響力の少なさがソーシャル・デジタルライブラリ全般の特徴としてみられるのであれば、それは本研究で提案した人物推薦手法が異なる分野に属するコミュニティを結びつける機会を提供することを意味する。また、第 2 章で述べたように、オンラインコミュニティは関心コミュニティを起点として、学習者コミュニティや実践コミュニティなどの様々な種類のオンラインコミュニティへと段階的に発展する。形状類似スコアが有効な実験協力者の中には、関心の幅が狭い NDC ツリーを有することを観察した。関心の幅の狭さは、ある特定の主題について強い志向性を持つことに繋がると考えられる。2.2 において、関心コミュニティの構成員の間で志向性と親密さが高まることで、学習者コミュニティや実践コミュニティへと発展する Henri のモデルについて述べた。Henri の理論に従えば、関心の幅が狭い実験協力者が関心コミュニティを形成した場合、志向性が高まりやすいことから、学習者コミュニティなど他のオンラインコミュニティへと発展することが考えられる。もし関心の幅とオンラインコミュニティの発展に関係があるならば、ソーシャル・デジタルライブラリのユーザが望むコミュニティの様態が多様であるときに、形状類似スコアがユーザひとりひとりのニーズに

従って人物推薦を行うことができる。すなわち、形状類似スコアの調節によって、関心の幅の狭いユーザには学習者コミュニティや実践コミュニティの形成支援を重視し、その一方で、関心の幅の広いユーザには関心コミュニティの形成・強化の支援を重点化するというように、ユーザごとにコミュニティ形成支援のプログラムを変更するといった手段をソーシャル・デジタルライブラリ上で提供することも考えられる。

したがって、形状類似スコアが示した、ユーザ間の関心の差異がコミュニケーションやコミュニティ形成に与える影響をより長期的な視点で研究することは、ソーシャル・デジタルライブラリのコミュニティ形成機能を発達させる手がかりとなる可能性がある。

第6章

結論

本論文では、ソーシャル・デジタルライブラリにおけるコミュニティ形成を目的とした人物推薦手法の確立を目指している。そこで、本論文では、NDC ツリープロファイリングを用いた人物推薦手法を提案した。NDC ツリープロファイリングとは、ユーザの資料利用履歴から日本十進分類法 (NDC) 番号を抽出し、ユーザの資料利用傾向を表現するツリー状のユーザプロファイルを作成する手法である。このユーザプロファイルを NDC ツリープロファイルと呼ぶ。そして、ユーザの NDC ツリー同士を、関心の共通項を反映した共通ラベルスコアと、関心の差異を反映した形状類似スコアという 2 つの尺度を用いて比較し、資料利用傾向の類似性を測定する。

本論文では、提案手法の有効性を検証するために、図書館情報メディア研究科の学生に対して評価実験を行った。その結果、ランダムに実験協力者を推薦するアルゴリズムと比較して、提案手法は高い精度・再現率・F 値を示した。このことから、提案手法は人物推薦として意義のある手法であることが明らかになった。

さらに、主題の階層性を考慮しない推薦方式よりも全体として再現率・精度が高いことが示され、パラメータの最適化において w_2 を重視することで結果が最適化されることが

示された。このことから、階層性を考慮した主題の共通性が人物推薦に良い結果をもたらすという本研究の仮説が検証された。

次に、実験協力者の資料利用履歴の大半を占める小説が与える影響を確認するために、小説を省いたデータで評価実験を行ったところ、重要視する主題の粒度を変更することで、小説を含んだ場合と同等の性能を達成できた。このことから、利用する資料利用履歴の傾向に応じて、主題の粒度の調節が可能である本手法の有用性が示されたと言える。

評価実験の結果から、関心の差異を表わす形状類似スコアについては、考慮することによって精度が上がるのが明らかになったが、その影響はわずかであった。ただし、形状類似スコアの有効性を詳細に検証するために、個別の実験協力者を見たところ、形状類似スコアを重視することで、より精度が上がる実験協力者もいることが明らかになった。

さらに、形状類似スコアが有効に働いた実験協力者は、関心の幅が狭い傾向を有していることが観察された。対照的に、形状類似スコアが有効に働いていない実験協力者は、関心の幅が広い傾向を有していることが観察された。このことから、実験協力者の関心の広がりや形状類似スコアの影響との間になんらかの関係性があることが示唆された。

本論文で明らかにした以上の結果は、限定的な実験環境下から導かれたものであり、今後の研究課題として次の2点が挙げられる。まず、実験条件の改善が挙げられる。提案手法の有効性を検証のためには、実験協力者数が十分大きく、様々な資料利用傾向を持つ実験協力者の集合に対して評価し、詳細なアンケートやインタビューを付随して行うなど、長期的かつ大規模な実験の実施が必要である。そのうえで、形状類似スコアが影響を与える場合と与えない場合を比較した上で、算出式の更なる改善が必要である。以上の課題を解決することで、ソーシャル・デジタルライブラリ上で実際にコミュニティを形成するまでの包括的な支援機能の開発へ研究が発展していくことが考えられる。

謝辞

筑波大学図書館情報メディア系の佐藤哲司教授には、本研究を進めるにあたりご指導と激励を賜りました。研究テーマに関する具体的な指導だけではなく、博士論文とはなにか、研究とはなにかといった心構えについてもご教示いただきました。こうして博士論文を書き上げることのできた今、ご教示いただいた教えのひとつひとつを忘れず精進してまいります。心から感謝の意を表します。

筑波大学図書館情報メディア系の緑川信之教授には、ご専門である分類法を中心として本研究についてご指導を賜りました。時には私自身には思い及ばない本研究の可能性についてご指摘いただき、本論文の意義を考える上で大変参考とさせていただきました。ご指導を参考にして、今後も分類法の可能性を伸ばしていけるよう尽力したい所存です。心から感謝いたします。

筑波大学図書館情報メディア系の宇陀則彦准教授には、デジタルライブラリの視点から本研究についてご指導を賜りました。図書館情報学における本論文の意義について深く考察するうえで、大変参考とさせていただきました。研究生活の中で何度も挫折しそうになった私を叱咤激励していただき、私に活力を与えてくださったご恩は忘れません。深く感謝いたします。

筑波大学図書館情報メディア系の中山伸一教授には、本論文のアルゴリズムおよび評価実験の記述を中心としてご指導を賜りました。本博士論文が他の研究者の方々の参考にな

るよう、誠実な記述を心がけるようご指導いただいたことは、今後の学術に対する取り組みを考える上で大変貴重な経験であったと考えております。ご指導頂いたことを深く感謝いたします。

放送大学教育支援センターの三輪眞木子教授には、研究に関して多大なるご指導をいただきました。特に考察の執筆にあたって必要な分析について様々なご指導を賜りました。唐突なご指導のお願いにもかかわらず、真摯に私の研究について耳を傾けていただき、時に厳しく時に温かい言葉を賜ったことを深く感謝いたします。

筑波大学図書館情報メディア系の松村敦先生には、研究についての基礎的な事柄から具体的な方策について長期にわたって丁寧に教えていただきました。様々な方向に迷走してしまいそうな私をさりげなく正しい方向へ導いてくださる先生のお力がなければ、ここまで辿り着くことはできませんでした。また、研究内容そのものだけでなく、研究と同時に健やかな生活を送るための日々の過ごし方や、学生への研究指導法について、先生ご自身の行動でもって示していただいたことは、これからの人生を充実させるための貴重な経験となりました。心からお礼申し上げます。

友人であり、本博士論文の核となる論文の共著者である小野永貴君には、本論文の研究のきっかけとなるソフトウェア開発プロジェクト Project Shizuku における活動をはじめとして、大学入学から現在にいたるまでの長期間に渡り様々な活動を共にし、その過程で研究に関する刺激と支援をいただきました。心からお礼申し上げるとともに、小野永貴君の今後の人生が公私ともにより豊かなものとなることを願っています。

Mac Book Air, LaTeX, Vim, Git, Ruby, R は、本論文を書き上げるうえで必要不可欠なツール群でした。また、WWW の存在なくしては私の研究を考えることはできません。これら偉大なテクノロジーを生み出した開発者の皆様に感謝いたします。

最後に、これまで私をあたたく応援してくれた家族に心から感謝します。

参考文献

- [1] Christine Borgman. What Are Digital Libraries? Competing Visions. *Information Processing & Management*, 1999, Vol. 35, No. 3, pp. 227–243.
- [2] Mark S. Ackerman. Providing Social Interaction in the Digital Library. *Digital Libraries'94*, 1994, pp. 198–201.
- [3] Marcos A. Gonçalves, Edward A. Fox, Layne T. Watson, and Neill A. Kipp. Streams, Structures, Spaces, Scenarios, Societies (5S): A Formal Model for Digital Libraries. *ACM Transactions on Information Systems*, 2004, Vol. 22, No. 2, pp. 270–312.
- [4] Maristella Agosti, M. Dobрева, V. Katifori, H. Schuldt, Leonardo Candela, Donatella Castelli, and Y. Ioannidis. The DELOS Digital Library Reference Model Version 0.98. *DELOS Network of Excellence on Digital Libraries*. DELOS Network of Excellence on Digital Libraries. http://www.delos.info/files/pdf/ReferenceModel/DELOS_DLReferenceModel_0.98.pdf (accessed 2014-04-04).
- [5] Michael E. Casey and Laura C. Savastinuk. Library 2.0. *Library Journal*, 2006, Vol. 131, No. 14, pp. 40–42.
- [6] Eric Brangier, Jérôme Dinet, and Laurent Eilrich. The 7 basic functions of a digital library - analysis of focus groups about the usefulness of a thematic digital

- library on the history of european integration. *Proceedings of the Symposium on Human Interface 2009 on Conferenceuniversal Access in Human-Computer Interaction*, 2009, pp. 345–354.
- [7] Adam Worrall. Social Digital Libraries : Their Roles within and Across Social Worlds, Information Worlds, and Communities. *Proceedings of the Doctoral Consortium, JCDL 2013, the 13th ACM/IEEE Join Conference on Digital Libraries*, 2013, pp. 55–62.
- [8] Susan L. Star and James R. Griesemer. Institutional Ecology, ‘Translations’ and Boundary Objects: Amateurs and Professionals in Berkeley’s Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, 1989, Vol. 19, No. 3, pp. 387–420.
- [9] Dimitris Gavrilis, Constantia Kakali, and Christos Papatheodorou. Enhancing Library Services with Web 2.0 Functionalities. *Research and Advanced Technology for Digital Libraries*, Lecture Notes in Computer Science, 2008, Vol.5173, pp.148-159.
- [10] 井上創造, 堀優子, 池田大輔. SNS(social networking service) における信頼と図書館における応用. 情報処理学会研究報告, 2007-FI-88, November 2007, Vol. 2007, No. 109, pp. 31–42.
- [11] Scott Jeffries. Social Cataloging Tools: A Comparison and Application for Librarians. *Library Hi Tech News*, 2008, Vol. 25, No. 10, pp. 1–4.
- [12] Louise F. Spiteri. The Impact of Social Cataloging Sites on the Construction of Bibliographic Records in the Public Library Catalog. *Cataloging & Classification Quarterly*, 2009, Vol. 47, No. 1, pp. 52–73.

-
- [13] 増井俊之. 本棚通信: 控え目なグループコミュニケーション. *インタラクシオン 2005* 論文集, 2005, pp. 135–142.
- [14] France Henri and Béatrice Pudelko. Understanding and Analysing Activity and Learning in Virtual Communities. *Journal of Computer Assisted Learning*, 2003, Vol. 19, No. 4, pp. 474–487.
- [15] Gerhard Fischer. Communities of Interest: Learning through the Interaction of Multiple Knowledge Systems. *Proceedings of the 24th IRIS Conference*, 2001, pp. 1–13.
- [16] Jean Lave and Etienne Wenger. Situated learning, legitimate peripheral participation. *Cambridge University Press*, 1991. 138p. (佐伯胖 訳. 状況に埋め込まれた学習: 正統的周辺参加. 産業図書, 2刷, 1995. 204p.).
- [17] Henry Kautz, Bart Selman, and Mehul Shah. Referral Web: Combining Social Networks and Collaborative Filtering. *Communications of the ACM*, March 1997, Vol. 40, No. 3, pp. 63–65.
- [18] Loren Terveen and David W. McDonald. Social Matching: A framework and Research Agenda. *ACM Transactions on Computer-Human Interaction*, September 2005, Vol. 12, No. 3, pp. 401–434.
- [19] Danah Boyd and Nicole B. Ellison. Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 2007, Vol. 13, No. 1, pp. 210–230.
- [20] Jilin Chen, Werner Geyer, Casey Dugan, Michael Muller, and Ido Guy. Make new friends, but keep the old: Recommending people on social networking sites. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*,

- 2009, pp. 201–210.
- [21] Luiz Pizzato, Tomasz Rej, Joshua Akehurst, Irena Koprinska, Kalina Yacef, and Judy Kay. Recommending People to People: the Nature of Reciprocal Recommenders with a Case Study in Online Dating. *User Modeling and User-Adapted Interaction*, 2013, Vol. 23, No. 5, pp. 447–488.
- [22] Kaizhong Zhang and Dennis Shasha. Simple fast algorithms for the editing distance between trees and related problems. *SIAM Journal on Computing*, 1989, Vol. 18, No. 6, pp. 1245–1262.
- [23] Mao Tsunekawa, Haruki Ono, Kyoji Konishi, Keita Tsuji, Atsushi Matsumura, and Norihiko Uda. Shizuku2.0: Cooperative Reading Support System. *Proceedings of the ASIA-PACIFIC CONFERENCE ON LIBRARY & INFORMATION EDUCATION & PRACTICE 2011 (A-LIEP 2011) Issues, Challenges and Opportunities*, 2011, pp. 539–547.

全研究業績のリスト

学位論文

1. 常川真央. 相互評価を支援する e ポートフォリオシステムの研究. 筑波大学, 2009, 卒業論文.
2. Mao TSUNEKAWA. Shizuku2.0: Community Oriented Reading Support System. 筑波大学, 2011, 修士論文.

論文 (査読付き)

1. 常川真央, 松村敦, 宇陀則彦. 日本十進分類法を用いた類似読者発見手法. 情報メディア研究, 2013, Vol.12, No.1, pp.42-51.

国際会議論文

1. Mao Tsunekawa, Haruki Ono, Kyoji Konishi, Keita Tsuji, Atsushi Matsumura, Norihiko Uda, Shizuku2.0: Cooperative Reading Support System, Proceedings of the ASIA-PACIFIC CONFERENCE ON LIBRARY & INFORMATION EDUCATION & PRACTICE 2011 (A-LIEP 2011) Issues, Challenges and Oppor-

tunities. 2011, pp.539-547.

論文 (査読なし)

1. 常川真央, 小野永貴, 松村敦, 宇陀則彦. 学習ノウハウの共有を支援するコミュニティ指向型図書館システム. 2010 年度人工知能学会全国大会 (第 24 回) 論文集. 2010. <https://kaigi.org/jsai/webprogram/2010/pdf/456.pdf>
2. 常川真央, 松村敦, 宇陀則彦. 情報検索教育を支援する「つまずき」発見ツールの提案. 日本教育工学会第 25 回全国大会講演論文集, 2009, pp.528-529.
3. 常川真央, 小野永貴, 安西慧, 矢ヶ部光. 利用者のつながりを創り出すコミュニティ指向型図書館システム (テーマ:学生チャレンジ特集). 情報処理学会研究報告 DD. 2008, Vol.34, pp.1-6.

雑誌記事

1. 小野永貴,常川真央. Web 時代にあるべき未来の図書館サービスの胎動, 貸出履歴の議論を超えた Shizuku2.0 の実現へ. 情報管理. 2010, Vol.53, No.4, pp.185-197.
2. 常川真央. 利用者サービスを指向した図書館システムの実現手法, Shizuku2.0 および Code4Lib Japan の活動を通して (図書館システムの可能性). 薬学図書館. 2010, Vol.55, No.4, pp.270-276.

著書

1. 岸田和明, 常川真央. 第5章 ウェブ情報資源の管理とアクセス. 情報資源の組織化と提供. 根本 彰, 岸田 和明. 初版, 東京大学出版会, 2013, pp.119-151. (シリーズ図書館情報学 2). ISBN:9784130034920.