

## 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

著者	常川 真央
発行年	2014
学位授与大学	筑波大学 (University of Tsukuba)
学位授与年度	2014
報告番号	12102甲第7091号
URL	<a href="http://hdl.handle.net/2241/00124075">http://hdl.handle.net/2241/00124075</a>

氏名(本籍)	常川 真央(東京都)
学位の種類	博士(図書館情報学)
学位記番号	博甲第7091号
学位授与年月日	平成26年5月31日
学位授与の要件	学位規則第4条第1項該当
審査研究科	図書館情報メディア研究科
学位論文題目	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

主査	筑波大学	教授	博士(工学)	佐藤 哲司
副査	筑波大学	教授	工学修士・文学修士	緑川 信之
副査	筑波大学	准教授	博士(工学)	宇陀 則彦
副査	筑波大学	教授	博士(学術)	中山 伸一
副査	放送大学	教授	Ph.D. in Information Transfer	三輪 眞木子

## 論文の要旨

本博士論文の目的は、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である。デジタルライブラリは、社会的文脈が異なる多様なユーザが利用することを前提としており、コミュニティを形成する媒介物として機能すると言われている。本論文では、オンライン上でコミュニティを形成する機会を提供する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、ソーシャル・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、ユーザプロフィールが類似した他のユーザを発見・推薦するための手法を提案している。本論文の構成は以下のとおりである。

第1章では、近年、デジタルライブラリがコミュニティ形成に寄与するとして注目されており、その機能に着目した新たなモデルとしてソーシャル・デジタルライブラリが提唱されたこと、ソーシャル・デジタルライブラリでは、コミュニティ形成を促進する組織、物体、仕組みを表す総称である「バウンダリオブジェクト」が重要であること等を説明したうえで、バウンダリオブジェクトの具体的実装を行うことで新たな人物推薦手法を提案すると述べている。

第2章では、本研究に関連するソーシャル・デジタルライブラリ、オンラインコミュニティ、人物推薦システムの3つの研究領域について先行研究を整理している。一つ目のソーシャル・デジタルライブラリについては、ソーシャルOPAC、ソーシャルカタログ・サイト、読書系SNSの3つの取り組み事例を紹介し、ソーシャル・デジタルライブラリのタイプ分けを試みている。二つ目のオンラインコミュニティについては、コミュニティの様態を関心コミュニティ、目的志向コミュニティ、学習者コミュニティ、実践コミュニティに区分し、関心コミュニティから実践コミュニテ

ィへと段階的に発展するものであると説明している。なかでも関心コミュニティがソーシャル・デジタルライブラリにおいては重要であり、資料利用履歴からいかに共通の関心を抽出するかが鍵であると述べている。三つ目の人物推薦システムについては、ユーザ間の関係性に基づく Friend-of-Friend アプローチとユーザが利用する資料の類似性に基づく Content-Matching アプローチの2種類があると説明し、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としている。以上のことから、ソーシャル・デジタルライブラリには Content-Matching アプローチの人物推薦手法が適しているとまとめている。

第3章では、本論文の人物推薦手法の核となる NDC ツリープロファイリングについて説明している。ここで、NDC とは日本十進分類法 (Nippon Decimal Classification) のことで階層構造を持った分類法である。提案する NDC ツリープロファイリングは、ユーザの関心を NDC のツリー構造でプロファイルする手法である。ソーシャル・デジタルライブラリでは、資料利用履歴をユーザの関心を推測する手がかりとすることができ、個々のユーザに対して、そのユーザの資料利用履歴に含まれる NDC 番号を NDC ツリープロファイルとして表現することにより、ユーザの関心を定量化できるとしている。

第4章では、NDC ツリープロファイルに基づく人物推薦アルゴリズムを提案している。そこでは、NDC ツリープロファイル同士の類似性を測る尺度として、形状類似スコア、共通ラベルスコア、コメント率の3種類を設定し、人物推薦を実現する類似度算出アルゴリズム NDCTREE1 を提案している。これら3種類の尺度を用いた利用者実験の結果、実験の範囲内ではあるがコメント率は性能向上に寄与しなかったことから、コメント率を除いた2種類の尺度を用いて改良したアルゴリズム NDCTREE2 を提案している。NDCTREE2 では、共通ラベルスコアに NDC の階層ごとの重みパラメータ  $w_1$ 、 $w_2$ 、 $w_3$  を設定するとともに、形状類似スコアと共通ラベルスコアの重みを調整するためのパラメータ  $u$  を設定している。

提案手法の有効性を検証するために、既存手法としてランダムにユーザを推薦する RANDOM 法、資料利用履歴をベクトル表現してコサイン距離で類似度を測る VECTOR 法、および、利用した資料の一致度を測る BOOK 法を実装し、提案手法と比較評価を行った結果、既存の手法より高い性能で人物を推薦できるパラメータの組み合わせが存在することを明らかにしている。ただし、実験で使用したユーザの資料利用履歴には小説が多く、データに偏りがあることも示している。

第5章では、データの偏りが結果に与える影響を検証するため、小説のデータを抜いた評価実験を行っている。その結果、階層構造のパラメータを調整することで小説を含めた場合と同等の性能を達成できるとしている。また、形状類似スコアの重みが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ことも述べている。一方で、形状類似スコアが有効に機能している人物推薦のケースも確認できている。

第6章では、今回の実験データの範囲内に限定されとはしながらも、提案する手法が既存の手法より優れた性能で人物を推薦できると結論づけている。その上で、本博士論文で提案した手法の特徴を次のように述べている。まず、ユーザ間の類似度を資料の共通性ではなく主題の共通性としたことで、ユーザ間で一致する資料が無い、あるいは極めて限定的であった場合でも、主題に共通性があればユーザを推薦できること、ユーザの関心に図書分類番号を用いたことで、これまで蓄積されてきた図書分類に関する知見を活かせること、更には、ユーザプロファイルを階層構造としたことで、階層ごとに設けたパラメータを調整することで重視する階層を柔軟に制御できることで

ある。

最後に、今後の課題と展望について述べ、論文を締めくくっている。

## 審 査 の 要 旨

### 【批評】

資料提供の機能が重視されてきたデジタルライブラリにおいて、資料そのものあるいは資料の提供には利用者同士を仲介する役割があることが指摘され、資料を媒介としたソーシャル機能もまたデジタルライブラリの重要な機能であるという認識が広がっている。このような背景の下、デジタルライブラリのコミュニティ形成を支援する人物推薦アルゴリズムを提案する本研究は、時宜を得た取り組みであり、研究の動機として十分理解できる。

本研究の基本的なアイデアは、資料に付与された NDC 番号がユーザの関心を表すユーザプロフィールに使えると考えたこと、また、NDC の階層に応じてユーザプロフィールをツリー構造で表現したことである。これにより、形状類似スコアと共通ラベルスコアというツリー構造の類似性に基づく人物推薦アルゴリズムの提案へとアイデアが展開されている。論文では、提案するアルゴリズムが実際に共通の関心をもつ人物を推薦できるかを利用者実験によって確認しており、アイデアの新規性、結果の信頼性ともに、学位論文としての要件を十分に満たしていると言える。以下、各章毎に批評を行う。

第 1 章は、利用者同士のコミュニケーションを重視したソーシャル・デジタルライブラリの意義や課題について述べた章である。ソーシャル・デジタルライブラリでは、共通の関心を持つユーザ同士がコミュニティを形成しやすいと考えられる。そのため、人物推薦においては、興味の類似性を測る手法が必要であるとしている。このような課題の設定は妥当なものであり、この課題をソーシャル・デジタルライブラリ内の資料に付与されたデータを用いて解決しようとする本研究の意義は十分に認められる。

第 2 章は、上記の課題に関する先行研究を整理した章である。ソーシャル・デジタルライブラリ、オンラインコミュニティ、人物推薦システムの 3 つの研究領域から関連研究を整理していることは、本研究の位置づけを明確にするという意味で妥当な観点である。ソーシャル・デジタルライブラリについての言及は、本研究が対象とする研究領域の全貌を把握するのに不可欠な要点をコンパクトにまとめている。オンラインコミュニティの関連研究は、コミュニティの区分と段階を述べており、ソーシャル・デジタルライブラリとの関連が理解できるようになっている。一方で、人物推薦は、図書館情報学以外の情報学分野でも研究開発が盛んな領域であり、サーベイとしてはやや物足りなさを感じる。

第 3 章は、本論文で提案する人物推薦手法のアイデアの核となる NDC ツリープロファイリングについて詳述した章である。プロファイリングに必要な特徴データに資料利用履歴を使う理由、特に NDC 番号を使う妥当性について説明し、NDC ツリープロファイルの構造について説明している。資料利用履歴が読書の関心を表しているとしてユーザプロフィールに利用することは合理的であると判断できるが、議論が早期に NDC ツリーに収斂したことは、実装を確かなものとするには有効であろうが、提案手法の適用範囲・敷衍性という観点からは、他の分類体系での実装も視野に入れるなど、もう少し一般化した議論が欲しいところである。結果として提案手法は、NDC の体系に強く依存し

た手法となっているが、日本の図書館における NDC の普及率の高さを鑑みると着実な一歩として評価できる。

第 4 章は、NDC ツリープロファイルをもとにした人物推薦アルゴリズムおよび利用者実験について説明した章である。評価のための利用者実験は、実験参加者 37 名による NDCTREE1 アルゴリズムの評価と、実験参加者 24 名による NDCTREE2 アルゴリズムの評価を行っている。最初の実験では既存手法の一つである RANDOM より高い精度と再現率を示していたが、不要なパラメータの削除やパラメータ間の連結法の再考を行ってアルゴリズムに改良を加えた NDCTREE2 を提案し、2 回目の実験を行っている。その結果、従来法として実装した RANDOM、VECTOR、BOOK と比較して精度と再現率の向上を確認したとしている。しかしながら、結果を詳細にみると、10 ある第一階層の分類のうち、0 類から 8 類までは登録された書籍の冊数が最初の実験で 10 冊程度、2 回目の実験で 20 冊程度であり、本研究の人物推薦のアルゴリズムの有効性を検証するにはデータ数が少ないと言わざるを得ない。また、文学を表す 9 類にデータが偏っている点も提案手法の核となる形状類似スコアの有効性が十分に評価できなかつた要因となつていよう。理想状態における疑似データを用いた実験を行うなどの慎重さが求められるところである。

第 5 章は、第 4 章で実施した実験では定量的に示せなかつた結果を補うために、9 類のデータを含めた場合と除外した場合の比較分析によって、さらには、実験参加者ごとの NDC ツリープロファイルを分析することによって提案手法の妥当性を考察している章である。9 類に関する比較分析では、資料利用履歴データの傾向に応じて重要視する主題の粒度を調節することで、データの偏りは吸収できることを明らかにしている。実験参加者ごとの NDC ツリープロファイルの分析では、特徴的なパターンのデータを詳細に分析することで、形状類似スコアが有効に機能する可能性があることを明らかにしている。このことは、形状類似スコアに対する重みパラメータを調整することで、推薦の傾向を制御できることを示唆しているとも言える。

以上のことから、十分な数のデータがあれば本研究が提案する人物推薦アルゴリズムが有効に機能するとした本考察は説得力のあるものとなっており、具体的なアルゴリズムには少なからず改良の余地はあるものの、提案手法そのものの評価には信頼性があると言える。

第 6 章は、まとめの章であり、提案する人物推薦アルゴリズムが、第 1 章で設定した本論文の目的を達成していることを述べている。

以上を総合的に判断すると、本論文は図書館情報学の学位論文として十分な内容を含んでいると認められる。

### 【最終試験結果】

平成 26 年 5 月 2 日、図書館情報メディア研究科学位論文審査委員会において、審査委員全員出席のもと、本論文について著者に説明を求めた後、関連事項について質疑応答を行った。引き続き、「図書館情報メディア研究科博士後期課程（課程博士）の学位論文審査に関する内規」第 23 項第 3 号に基づく最終試験を行い、審議の結果、審査委員全員一致で合格と判定された。

### 【結論】

よって、本学位論文の著者は博士（図書館情報学）の学位を受けるに十分な資格を有するものと認められる。