

情報学分野を対象とした研究費獲得状況と図書館所蔵情報の 横断検索システムの開発とそのFDへの応用

安川 美智子（群馬大学）

1. はじめに

2019年新型コロナウイルス感染症(COVID-19)の影響により、2020年度の学会発表や大学の授業の多くがオンラインで配信されるようになり、また、配信内容が動画として保存されている場合もある。このような動画のライブ配信（同期型）やオンデマンド配信（非同期型）を視聴することは、大学における研究と教育の質を向上するためのファカルティ・ディベロップメント（以下、「FD」という。）に役立つと考えられる。しかし、動画の視聴には時間がかかるため、視聴に値する学会発表や模範的な授業を行っている優秀な教員を効率よく発見し、その教員の属性情報（姓名、授業を担当している大学名および部局名、研究テーマや学会発表先、教科書・参考書、等）をデータベースから検索できることが望ましい。大学教員の能力開発のためのFDについては、大学の学士課程教育に期待される取り組みに関する答申において、以下のような指摘がされている[1]。

【指摘1】 高度な専門職である大学教員について、共通して求められる専門性が存在する一方、その多様な在り方も尊重されなければならない。

【指摘2】 教育と研究との相乗効果が発揮される教育内容・方法を模索していく必要がある。

【指摘3】 教員の人事・採用に当たっての業績評価について、研究面に偏することなく、教育面を一層重視する。評価の対象として、例えば、優れた教科書や教材の作成についても積極的に位置付ける。

以上のことを研究の背景として、本研究では大学における教員の能力開発を支援するためのデータベース検索手法を提案する。本研究で明らかにする研究上の問い(Research Question)は以下の2点である。

【RQ1】 研究者によるピア・レビュー審査において十分な研究遂行能力があると評価され、かつ、優れた教科書や参考書を執筆している教員をデータベースから検索することは可能だろうか。

【RQ2】 データベースから検索された教員（教育と研究との相乗効果を発揮していると思われる教員）の中から、他者を優越する優秀な教員を発見できるだろうか。

以下、本論文では、情報学分野を対象として、研究費獲得状況と図書館所蔵情報の横断的検索を行う検索システムを提案し、そのFDへの応用可能性について検討する。

2. 研究費獲得状況の検索

科学研究費助成事業データベース KAKEN[2]は、科学研究費助成事業（学術研究助成基金助成金／科学研究費補助金）（以下、「科研費」という。）の採択課題の概要と報告書のデータを収録している。このデータベースを用いることにより、科研費の採択課題の研究課題番号、研究課題名、審査区分/研究分野、助成金の配分額、研究代表者の情報（研究者番号、姓名、所属機関）、および、当該課題の研究成果の学会発表先などを調査できる。データベース KAKEN の XML データは随時、更新・修正が行われている。本研究では 2020 年 8 月 14 日～24 日の期間に収集したデータを使用した。収集したデータの中には研究課題番号が 942,477 件収録されており、審査区分/研究分野が設定された研究課題番号の数は 846,252 件、審査区分/研究分野に「情報」の文字列を含む研究課題番号の数は 26,105 件であった。また、助成金の配分額の未設定が 3,545 件、研究代表者の姓名の未設定が 43 件であり、これらについては検索対象外とした。審査区分/研究分野が設定された研究課題を研究代表者ごとに集約し、「採択された研究課題の件数」と「助成金の配分額の合計値」の 2 変数を 2 軸として作成した散布図を図 1 に示す。また、図 1 の研究代表者のデータから、以下の 2 つの条件を満たすデータを選択して、図 1 と同様に作成した散布図を図 2 に示す。

【検索条件 1】 科研費の採択課題が「情報学分野」で 1 件以上ある研究代表者

【検索条件 2】 大学図書館に所蔵の著書が「情報学分野」で 1 冊以上ある研究代表者

検索条件 1 は検索結果を審査区分/研究分野が「情報」であるものに限定する条件である。検索条件 2 は検索結果を「情報学分野」の著書の所蔵情報（3 章で述べる）がある教員に限定する条件である。図 1 と図 2 の散布図の各点は科研費の研究者番号がある各個人に対応している。図 1 と図 2 のどちらの場合も、科研費の「採択件数が最大値」の研究者と「配分額の総額が最大値」の研究者は、異なる 2 名の研究代表者である。「採択件数」と「配分額の総額」の両方の観点からデータベースの検索を行うことで、研究期間や予算規模の異なる多様な研究者を検索できる。

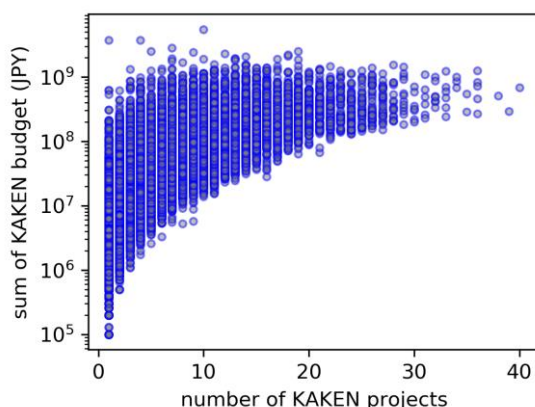


図 1 科研費データベース (全体)

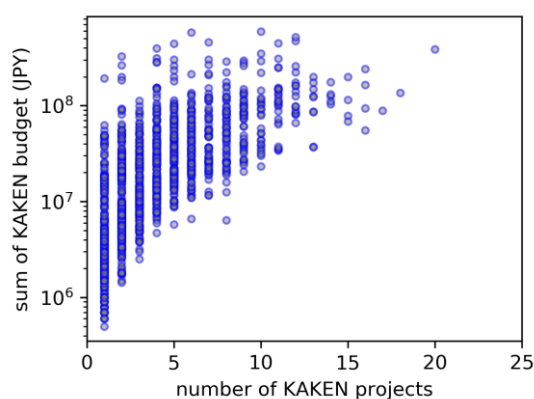


図 2 科研費データベース(情報学分野)

3. 図書館所蔵情報の検索

図書館所蔵データのデータベース CiNii Books は全国の大学図書館約 1,300 館が所蔵する約 1,177 万件(のべ 1 億 4 千万冊以上)の本を検索できる学術情報プラットフォームである[3]。CiNii Books を用いることにより、本の識別子(NCID)、タイトル、著者名、出版者、出版年、所蔵図書館、および、日本十進分類法(NDC)に基づく本の分類などを検索できる。本研究では、情報学分野の本を対象とするため、5 つの NDC の分類 007 (情報科学), 547.48 (データ通信), 548 (情報工学), 694 (産業・経営・事業に関する観点), 810 (日本語)のいずれかが付与された本の RDF データを収集した。CiNii Books の情報は日々、更新されている。本研究では、2020 年 8 月 29 日～31 日の期間に収集した RDF データ(79,654 件の NCID、43,830 人の著者、図書館所蔵のべ 5,211,386 冊)を使用した。収集したデータを分析し、本の著者が科研費データベース KAKEN の研究代表者として登録されているか否かを調べ、分類ごとに視覚化した棒グラフを図 3 に示す。横軸は本の分類(分析対象とした 5 つの NDC 分類)、縦軸は本の数(本の識別子 NCID の数)である。緑色とオレンジ色のグラフは、それぞれ、本の著者が KAKEN に登録されているか否かを表している。図より、CiNii Books の情報学分野の本の数に対して、著者が科研費の研究代表者である本の割合は比較的小さいことがわかる。収集したデータを著者ごとに集約し、「著者が執筆した本の NCID の数」と「著者が執筆した本の図書館所蔵数の和」の 2 変数を 2 軸として作成した散布図を図 4 に示す。図 4 のデータは以下の 2 つの検索条件を満たす本のデータである。

【検索条件 3】 科研費の採択課題が「情報学分野」で 1 件以上ある著者

【検索条件 4】 大学図書館に所蔵の著書が「情報学分野」で 1 冊以上ある著者

図 4 において、「本数が最大値」の著者と「本の図書館所蔵数の和が最大値」の著者は異なる 2 名の著者である。「本数」と「所蔵数の和」の両方の観点から本のデータを検索することで、本を多数執筆している著者であるか、あるいは、執筆した本の人気度の高い著者であるか、といった異なる観点からの多様な著者検索が可能となる。

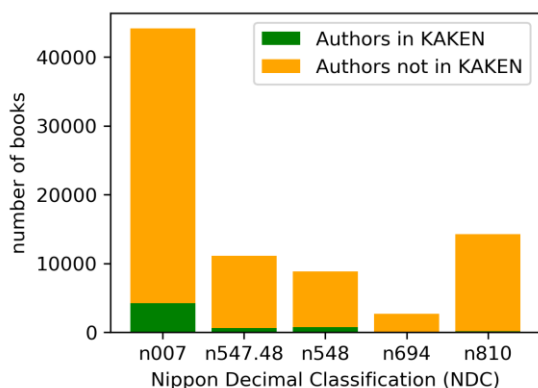


図 3 著者が科研費代表者であるか否か

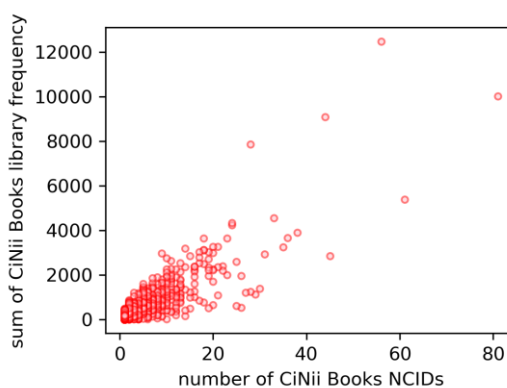


図 4 図書館所蔵情報(情報学分野)

4. 研究費獲得状況と図書館所蔵情報の横断検索

図 2 と図 4 の点の集合から、科研費獲得、および、本の執筆の両方の観点から他者を優越している教員（研究者）を発見するため、KAKEN と CiNii Books の横断検索を行った。「助成金の配分額の合計値」と「著者が執筆した本の図書館所蔵数の和」の 2 変数を 2 軸として作成した散布図を図 5 に示す。また、「採択された研究課題の件数」と「著者が執筆した本の NCID の数」の 2 変数を 2 軸として作成した散布図を図 6 に示す。図 5 には他のデータから離れたデータ（はずれ値）が 2 件ある。詳細を確認した結果、これら 2 件のデータに対応する研究代表者 2 名はトップレベルの研究業績と多数の本の執筆の実績がある研究者であることが確認できた。他のデータから外れているデータ（ユートピア個体）を取り除き、残った点の集合にスカイライン演算[4]を適用すると、点の集合の輪郭に位置する点、すなわち、他者に優越されない点の集合（パレート最適解の集合）を得ることができる。図 5 と図 6 のパレート最適解の集合を赤色でプロットした散布図を図 7 と図 8 に示す。得られた解の集合は、「科研費獲得」と「本の執筆」のいずれか一方の観点からの検索では他の点からの分離が困難な点であったが、両方の観点からの検索（提案法）により、発見することが可能となった。以上のことから、1 章で述べた RQ1 および RQ2 の「優秀な教員の発見」が可能であることが実際のデータを用いた実験により、確認できた。

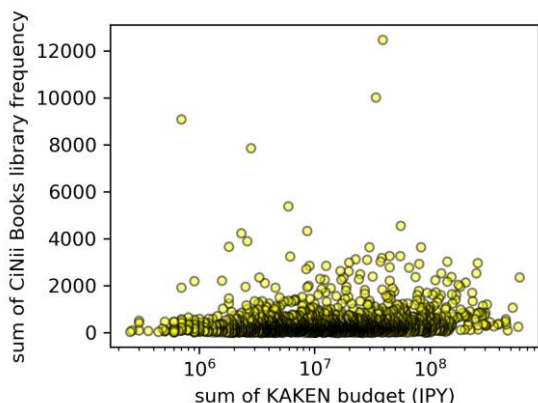


図 5 「科研費総額」と「著書の所蔵数」

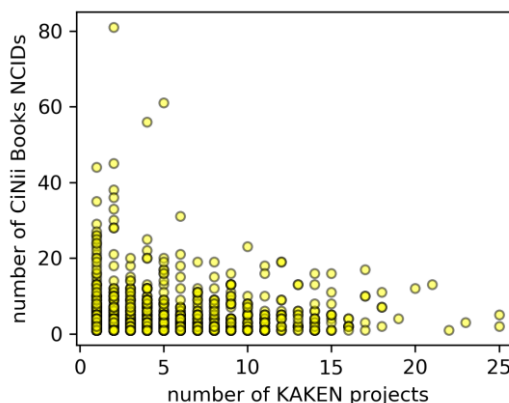


図 6 「科研費獲得件数」と「著書の数」

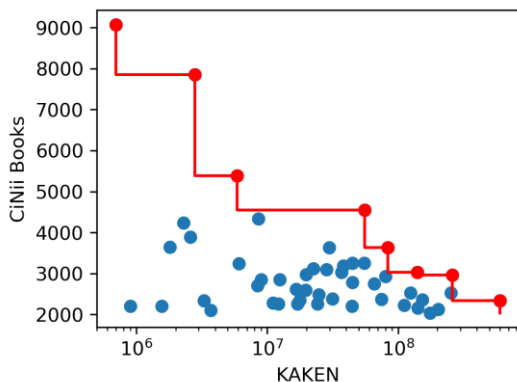


図 7 パレート最適解(図 5 から抽出)

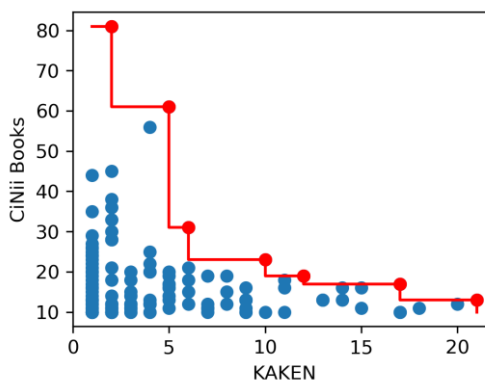


図 8 パレート最適解(図 6 から抽出)

5. 考察

群馬大学では教員の活動状況を点検・評価する教員評価を毎年実施している[5]。教員評価の領域としては、「教育」「研究」「社会貢献」及び「管理・運営」の4つの領域が定義されている。また、群馬大学では令和元年・2年度の組織目標の一つとして、FD活動等を通じて教職員の能力発展を推進することを掲げている。これまでに、「外部研究資金の獲得者が申請書の書き方のコツを学内研究者に指南する説明会」や「ベストティーチャー最優秀候補者による公開模擬授業」などのFD活動が実施されている[6]。このような教員全体に向けてのFDセミナーに加えて、本研究で提案するようなデータベース検索ツールを活用し、各教員が得意分野や適性に応じた能力開発を自発的に行うことができれば理想的である。具体的には、各教員が「教育」と「研究」の両方の観点から参考にしたい優秀な教員をデータベースから発見し、優秀な教員の属性情報（姓名、所属大学名と部局名、科研費の報告書、執筆した本の書誌情報、等）を手掛かり情報として、学会発表や授業の動画を検索・視聴することで、FDをさらに促進できると考えられる。

提案法の現状の問題点としては、大学教員の中には同姓同名の人がいるため、このような例外的なデータについては手作業でデータ整備していく必要があり、分析用のデータを準備する際に時間と手間がかかるという点である。本研究で収集したデータの中にも、姓名の対応付けが自動処理できないものが存在した。具体的には、情報学分野では、8,856名の研究者のうち289名(3.26%)に同姓同名の他者が存在していた。大規模なデータベースの不具合の特定と解消を行うデータ整備作業は、研究IRや教学IRのさまざまな実務において必要となる作業であるが、時間と手間がかかる作業であるため、教職員の誰かがいったん整備したデータを同じ大学内や国内の大学間で共有・再利用できる仕組みが構築されれば、日本におけるIR業務の効率化につながると考えられる。

6. おわりに

本発表では、大学教員の能力開発のためのFD活動を支援することを目的として新たなデータベース検索法を提案した。提案法により、教員の「研究」または「教育」のどちらか一方に偏るのではなく、教育と研究との相乗効果が発揮されるような能力向上を、教員がインターネット上の動画等の資源を活用して自発的に行えるように支援できる。今後の課題としては、より精度の高い分析を行えるようにするための研究者の姓名の「名寄せ」（同じ人の姓名の異表記の集約、同姓同名の人の区別、等）のためのデータ整備が必要であると考えている。

【謝辞】

本研究はJSPS科研費(JP18K11986)、および、統計数理研究所共同研究プログラム(2020-ISMCRP-0012)の助成を受けたものです。

【参考文献】

- [1] 学士課程教育の構築に向けて，平成20年12月24日，中央教育審議会
http://staff.gku.ac.jp/~soumu/data/1.toushin/H20_1224gakushi.pdf
- [2] KAKEN：科学研究費助成事業データベース（国立情報学研究所）
<https://kaken.nii.ac.jp/>
- [3] CiNii パンフレット，2018年4月，国立情報学研究所
https://support.nii.ac.jp/sites/default/files/cinii_pamphlet_web_2018.pdf
- [4] S. Borzsony, D. Kossmann and K. Stocker, "The Skyline operator," Proceedings 17th International Conference on Data Engineering, Heidelberg, Germany, 2001, pp. 421-430, doi: 10.1109/ICDE.2001.914855.
- [5] 群馬大学 教員評価
https://www.gunma-u.ac.jp/outline/out006/out006_001/g1784
- [6] 平成30事業年度に係る業務の実績に関する報告書，令和元年6月 群馬大学
<https://www.gunma-u.ac.jp/wp-content/uploads/2019/07/30houkokusho.pdf>