

コレクションの多様性に関する評価指標と適用可能性の検討

武田千佳*, 池内淳**

筑波大学

*s1811509@s.tsukuba.ac.jp **atsushi@slis.tsukuba.ac.jp

近年、図書館収集方針には多様な資料を収集しようとする姿勢が現れている一方、コレクション全体の多様性を評価しようとした研究は少ない。本研究では、生物多様性評価指標を基に定式化された学際性評価指標を援用し主題に基づき図書館の多様性を評価する評価指標を作成し、既存の異なり分類記号数等といった統計データによる多様性評価と比較して適用可能性を検討した。

Creating metrics for diversity of collection and considering their applicability

Chika TAKEDA*, Atsushi IKEUCHI**

Tsukuba University

1. はじめに

近年、図書館の多様性について様々な文脈で語られている。学術情報流通における「書誌多様性 (Bibliodiversity)」の形成を呼びかけるジュシュー宣言 (The Jussieu Call) [1]では、“様々なサービス、プラットフォーム、資金調達方法、評価指標の存在することが、多様なニーズに応えることを可能にする”と述べている。ここでは、多様性の存在によって、“人や価値観の新たなコラボレーションが生まれ、新しい発想やイノベーション (変革) のきっかけ”[2]となることが期待されている。アメリカ図書館協会 (ALA) 知的自由部は“図書館には多様な資料を収集する義務がある”[3]と述べている。日本においても多様な資料を収集しようとする姿勢が図書館収集方針に現れている。

このように資料や資料収集における多様性が重要視されている。これまで、特定のトピックや主題について網羅的に収集しているのかについて調査した事例[4][5]は存在する一方で、コレクション全体の多様性を評価しようとした研究は少ない。また、コレクション全体の多様性を測定するための尺度としては、異なり主題数や異なり分類記号数が挙げられるものの、それらは単に網羅性の尺度にすぎず、実際の図書館における収集方針を反

映した収集を行っているかについてまで踏み込んで測定するものではない。そこで、本研究では、コレクション全体の多様性を測定する指標を導入し、その適用可能性について検討することを目的とする。

2. 先行研究

図書館のコレクションの多様性を評価する既往事例としては、多様性に関する主題を持つ資料や、異なる見解を含む資料の所蔵調査が挙げられる。Ciszek ら (2010) [6]は、“コレクション分析の手法は定量的評価・定性的評価、利用者中心・コレクション中心という2つの軸によって4つのカテゴリに分類される”と述べる。

次に、他の分野の指標を基に評価指標を作成した研究として、学際性評価の研究がある。Rao (1982) [7]は、生物多様性評価指標である Simpson 指数に分野間の非類似性である disparity を掛け合わせた。計量書誌学ではこの尺度を“Rao-Stirling diversity indicator” (Rao-Stirling 多様性指標) と読んでいる。一方で Leydesdorff ら (2019) [8]は Rao-Stirling 多様性指標は学際性評価に適さないとして新しい測定指標 (DIV) を作成した。Stirling (2007) [9]と

Leinster ら (2012) [10]は多様性の概念に“variety (主題の多さ)”, “balance (蔵書バランス)”, “disparity (主題間非類似性)”という3つの構成要素があることを指摘した。Simpson 指数は「主題の多さ」と「蔵書バランス」の2つを, Rao-Stirling 多様性指標と DIV は「主題の多さ」「蔵書バランス」「主題間非類似性」の3つを構成要素としてもつ。

3. 研究方法

本研究ではコレクションの多様性を主題の多様性によって測定することを試みる。主題を示す概念はさまざまなものがあるが,ここでは主題を「日本十進分類法に基づく分類記号」によって定義する。上述した「異なり主題数」や「異なり分類記号数」といった統計データと,生物多様性を測定するための評価指標のどちらが図書館におけるコレクションの多様性評価に適しているか検討するために,ここでは,「実測値」「指標値」「理想値」を定義する。このうち実測値と指標値は実際に多様性評価に用いる値であり,実測値は統計データ,指標値は評価指標のことを指す。理想値は実測値と指標値のうちどちらが多様性評価に適しているか検討を行う際の基準として用いる値で,図書館が目指す理想の収集の程度を表すと考えられる,図書館収集方針で規定されている文言から集計した値と定義する。本研究では,実測値よりも指標値のほうがコレクションの多様性を一意に評価できると仮定して研究を行った。

3.1 実測値の測定

実測値およびそれに基づき算出される指標値の入力データは国立国会図書館サーチをクローリングして収集した日本十進分類(以下, NDC)とその NDC が付与された資料の所蔵冊数の組である。実測値として用いるのは各図書館の蔵書冊数(以下, コレクション数), および各図書館が所蔵する資料に付与された分類記号の総数(以下, 異なり

分類記号数)の2つである。

データの収集は都道府県立図書館のうち, 国立国会図書館総合目録ネットワークにデータ提供館として加盟している計 53 館を対象とする。国立国会図書館サーチをクローリングすることによって, 各図書館の所蔵データを収集した。クエリを自動生成するプログラムを作成し, 2021年10月9日から10月24日にかけてクローリングを行った。

3.2 指標値の計算

指標値は「実測値を入力データとして学際性評価指標でコレクションを評価した値」と定義する。主題の多さ, 各主題について蔵書冊数のバランス, 主題の似ていない程度の3つの軸で多様性を測定する。主題数が多いほど, 蔵書冊数が均等なほど, 同じ主題数でも主題同士が似ていないほど, 多様であると評価する。

具体的な指標は Simpson 指数, Rao-Stirling 多様性指標, DIV の3つを用いる。「主題の多さ」「蔵書バランス」「主題間非類似性」の3つの軸は

Simpson 指数

… 主題の多さ + 蔵書バランス

$$D = 1 - \sum_{i=1}^n P_i^2$$

Rao-Stirling 多様性指標

… Simpson 指数 + 主題間非類似性

$$RS = \sum_{i,j} q_i q_j d_{ij}$$

DIV

… 主題の多さ + 蔵書バランス + 主題間非類似性

$$DIV = \frac{n_m}{N} \times (1 - Gini) \times \sum_{i,j,i \neq j} \frac{d_{ij}}{d}$$

図1: 式と対応する軸

図1の数式と同じ色で囲われた部分(モノクロ印刷の場合、点線の長さが同じ部分)で反映される。Simpson 指数は「主題の多さ」と「蔵書バランス」の2つを1つの項で反映する。

3.3 理想値の計算

実測値と指標値のどちらが多様性評価に適しているか検討する基準として、理想値を定義する。理想値は図書館が目指す理想の収集を表し、図書館収集方針で規定されている文言から集計した各値とする。具体的な収集の姿勢が示された選定基準を集計の対象とし、オンライン上で公開されている12館の選定基準すべてに集計を行った。「異なり分類記号数」は選定基準に記述されていればその分類記号を、記述されていなければ NDC 典拠データベースに登録されている分類記号をカウントする。「異なり主題数」は記述の主語にあたる部分を主題としてカウントする。「収集の程度の点数」は記述の述語および述語にかかる修飾語を収集の程度を表す文言とみなし、収集の程度に応じた7段階の点数を加算していく。点数化の対象となる文言は、「網羅的に」「積極的に」「広く(幅広く、体系的に)」「選択的に」「限定的に」「厳選」「収集しない」の7つを基本としてこの順に6~0点をつける。これらの文言以外の文言は、意味が類似する基準となる7つの文言と同じ点数とした。

4. 結果

4.1 実測値と指標値間の相関

実際の収集コレクションを表す数値として指標値は適しているのか検討するために、実測値と指標値の間を検定した結果、以下の通りとなった。

Simpson 指数はコレクション数、異なり分類記号数のいずれの実測値とも有意ではなかったが、Rao-Stirling 多様性指標と DIV は有意であった。有意となった Rao-Stirling 多様性指標と DIV の相関係数を比較した結果、コレクション数と異な

り分類記号数のいずれの場合でも Rao-Stirling 多様性指標のほうが高かった。これらのことから、Rao-Stirling 多様性指標が最も図書館の実際のコレクション収集結果の多様性を反映していると考えられる。

4.2 理想値と実測値間の相関及び理想値と指標値間の相関

実測値と指標値のうち収集方針に示す理想的な収集をしている図書館を判断するのに適しているのはどちらか検討するために、理想値と実測値、理想値と指標値でそれぞれ検定を行った。

実測値と指標値のどちらがより選定基準を反映しているか検定を行ったところ、指標値のほうが理想値と有意である場合が多かった。相関係数を比較した結果、理想値の異なり主題数、異なり分類記号数、点数のいずれと比較した場合でも実測値(コレクション数、異なり分類記号数)より指標値(Simpson 指数、Rao-Stirling 多様性指標、DIV)のほうが係数は高かった。

この結果から、選定基準に従って多様な資料を収集しているか評価する際には、実測値(統計データ)より指標値(多様性指標)のほうが適していると考えられる。

さらに、指標値のうち異なり主題数、異なり分類記号数、点数のいずれの場合でも DIV の相関係数が最も高かったため、DIV が指標のなかで最も選定基準に従って多様な資料を収集しているコレクションを反映した数値が算出できると考えられる。

5. 考察

実際の収集コレクションの多様性を評価するには Rao-Stirling 多様性指標、選定基準に従って多様な収集を行っているか評価するには DIV が適していると導出結果に違いがみられた要因として、以下の3つが考えられる。

①選定基準が策定される以前に収集された資

- 料も本研究の調査対象に含まれていること
- ②選定基準をどれだけ詳細に記述するかは図書館によって異なること
- ③本研究では一律に定めた「主題の似ていない程度」を表す距離は実際には異なっていること

6. おわりに（研究の限界と今後の課題）

本研究では、学際性評価指標を流用し主題に基づいて図書館の多様性を評価する指標を作成し、資料冊数や異なり分類記号数といった図書館がすでに収集している統計データによる多様性評価と比較して検討を行った。その結果、資料冊数および異なり分類記号数そのものをバラバラに集計し分析するより、要素がまとまった多様性指標のほうが収集方針に従った収集をしているか判断するうえで適していると考えられる。しかしながら、分類構造などの指標の計算に関する問題、図書館の多様性について定まった定義がなく比較できる既存研究がないことから、今後さまざまなデータや指標を用いて研究を行い、その結果を総合的に判断する必要があると考える。

7. 注・文献

- [1] Shearer, Kathleen et al. 学術情報流通における「書誌多様性」の形成に向けて —行動の呼びかけ—. 河合将志ほか訳. 国立情報学研究所 オープンサイエンス基盤研究センター. 2020, <https://doi.org/10.20736/00001276>, (参照 2021-12-24).
- [2] “なぜ SDGs で「多様性」が大事なの?”. NHK. 2021-06-18. https://www.nhk.or.jp/campaign/mirai17/ki_ji_diversity.html, (参照 2021-12-24).
- [3] Office for Intellectual Freedom of the American Library Association (2006), "Diversity in collection development: an interpretation of the Library Bill of Rights", Intellectual Freedom Manual, ALA Press, Chicago, IL, pp. 117-8.
- [4] Wood, Colleen. Counting The Collection: Diversity Audits Assess Representation In Library Materials, But Most Examples Focus On Children's Books. One Librarian Shares Her Methods And Challenges In Tackling Adult Biographies. *Library Journal*. 2021, Vol. 146, No. 5, p. 26-.
- [5] Blakemore, Megan; Pattee, Amy. Accounting for Diversity: Measuring Change in the Proportion of African American Teen Literature in the Senior High Core Collection. *Journal of Research on Libraries and Young Adults*. 2021, Vol. 12, No. 1, p. 1-24.
- [6] Ciszek, Matthew P; Young, Courtney L. Diversity collection assessment in large academic libraries. *Collection Building*. 2010, Vo. 29, No. 4, p. 154-161.
- [7] Rao, C. R. Diversity: Its measurement, decomposition, apportionment and analysis. *The Indian Journal of Statistics, Series A*. 1982, Vol. 44, No. 1, p. 1-22.
- [8] Loet Leydesdorff; Caroline S.Wagner; Lutz Bornmann. Interdisciplinarity as diversity in citation patterns among journals: Rao-Stirling diversity, relative variety, and the Gini coefficient. *Journal of Informetrics*. 2019, Vol. 13, Issue 1, p. 255-269.
- [9] Stirling, A. A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society*. 2007, Vol. 4, No. 15, p. 707-719.
- [10] Tom Leinster; Christina A Cobbold. Measuring diversity: The importance of species similarity. *Ecology*. 2012, Vol. 93, No. 3, p. 477-489.