

大学図書館の所蔵情報に基づくマンガの類型化

江藤正己(学習院女子大学)*
大谷康晴(日本女子大学)

安形輝(亜細亜大学)
安形麻理(慶應義塾大学)
*masaki.eto@gakushuin.ac.jp

杉江典子(東洋大学)
橋詰秋子(慶應義塾大学大学院)

1. はじめに

図書館とマンガに関わる課題の一つとして、組織化の問題がある。たとえば分類の場合、小説と同様に¹⁾、多くのマンガに対して同一の分類記号が割り当てられることが多く、十分に組織化できていないことが指摘できる。アメリカでは、図書館員の視点からマンガの組織化を扱った文献が散見されるが、その多くは実践報告やノウハウ共有を意図したもので、組織化に関する学術的な研究はあまりみられない²⁾。日本においては、マンガのメタデータや分類を扱った学術研究がいくつか存在するが、セマンティックウェブなどの新たな情報技術の導入を主眼とした内容が多い³⁾。

本研究では、より実証的なマンガの組織化手法として、大学図書館の所蔵情報を用いることを提案する。ある大学図書館にあるマンガが所蔵されるか否かは大学図書館(員)によるマンガの評価結果といえるため、このデータを適切に利用できれば、大学図書館(員)のマンガに対する評価結果の集積に基づいて、マンガを組織化することができる。特に、大学図書館では、公共図書館よりもサービス対象が限定されるため、その専門性を反映した所蔵データとなっていることが考えられる。

近年、データベースやそれに付随する API が一般化してきており、図書館におけるマンガの所蔵状況を全国的に調査することが容易になってきている。たとえば、小山らは大学図書館の⁴⁾、筆者らは公立図書館の⁵⁾マンガの所蔵状況を調査している。多数の図書館の所蔵データに基づく、マンガの組織化は近年のデータベースの整備により展望が開けた新たな方法といえる。

より具体的には、一つの図書館と一緒に所蔵されているマンガ同士には何らかの関連性があるという共起に基づく類似度を設定し、それを用いてマンガのクラスタリングをおこなう。著者らは、既に公共図書館の所蔵データに対して、この手法の適用を試みており、マンガの組織化における有用性を確認している⁶⁾。本稿の目的は、大学図書館の所蔵データを用いたマンガの組織化の可能性を明らかにするため、

クラスタリング手法を用いたマンガの類型化をおこなうことである。

2. 大学図書館のマンガ所蔵データの収集

2.1 対象資料

データの収集対象としたマンガは、国立国会図書館所蔵の以下の条件を満たす図書で、127,563 タイトルになる。

- ISBN が付与されている
- NDLC の記号が Y84 から始まる、あるいは国会図書館の請求記号が Y84 から始まる
- 2013 年 12 月末までに出版されている (ISBN のあるタイトルのため 1980 年頃からが対象)

この収集対象資料のメタデータについては国会図書館サーチ API により入手した。

さらに、今回の調査では、シリーズ物については第一巻を所蔵した図書館はそれ以後も原則的には所蔵するであろうことを仮定し、収集対象を単巻もの、ないし、シリーズものの第一巻に限定した。具体的には dcndl:volume に“0”, “1”, “第 1 巻”, “1 巻”, “第 1 集”, “v.1”, “vol.1”, “volume 1”, “上巻”が含まれるタイトルを収集の対象とした。

2.2 収集結果

データ収集の対象館は、CiNii Books で調査可能な全国の大学図書館を中心とした図書館(2016 年末 1,321 館)である⁷⁾。2016 年 8 月に、収集対象資料の ISBN を検索クエリとして CiNii Books API を用いて ISBN から NCID (NII 書誌 ID) を検索した。NCID では多巻ものは一つにまとめられるため、改めて NCID で所蔵データを検索したのちに、所蔵データを ISBN 単位に戻し、国会図書館のデータと名寄せをおこなった。

この結果、3,224 タイトルが、1 館以上の対象図書館で所蔵されていることが分かった。本稿は、初期的な研究であり、大学図書館の所蔵データの組織化における可能性をみるものであるため、10 館以上に所蔵されている 712 タイトルをクラスタリングの対象とした。

3. クラスタリング

3.1 マンガ同士の類似度距離の算出方法

マンガ同士の類似度距離は Simpson 係数を用いて、次式により計算した。

$$Distance(X, Y) = 1 - \frac{|Set X \cap Set Y|}{\min(|X|, |Y|)}$$

この式において、X と Y はそれぞれ一つのタイトルを、 $\min(|X|, |Y|)$ は X と Y の所蔵館数のうちの小さい方の値を、 $|Set X \cap Set Y|$ は X と Y のどちらも所蔵している図書館の数を示している。Simpson 係数は値が大きいほど類似性が高いことを示す指標であるため、1から減算することで類似度距離に変換して、クラスタリングに用いる。

3.2 クラスタリングの手法

クラスタリング手法としては、一般的によく使われている、ウォード法に基づく階層クラスタリングを採用した。具体的には、統計パッケージである R-3.3.2 の `hclust` 関数を用いた。

3.3 階層クラスタリングの結果

図1が階層クラスタリングの結果を示すマンガタイトルのデンドログラムである。上位階層に八つのクラスタがあると解釈することができる。クラスタ 4, 2, 6, 7, 8 はそれぞれが独立している一方で、クラスタ 1, 3, 5 には大きなまとまりがあり、かつクラスタ 1・3 とクラスタ 5 に分かれていることが図1からよみとれる。

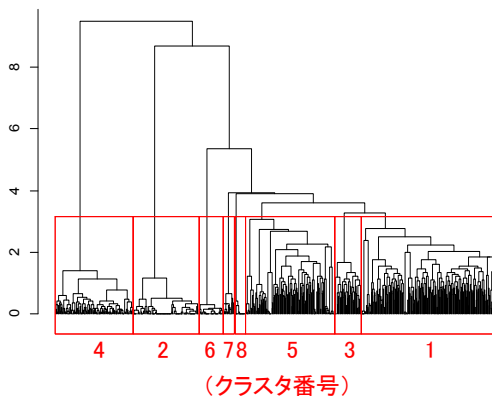


図1 マンガタイトルのデンドログラム

各クラスタに含まれるマンガのタイトルとクラスタに含まれるマンガのタイトル数を表1に示す。ここで挙げているタイトルは、当該のクラスタに含まれるマンガのうち、多くの図書館に所蔵されているもの、上位五つのタイトルを原則として選んでいる。また、クラスタのラベルは、クラスタに含まれるマンガの特徴を、これら上位五つを中心に、著者らが解釈しておこな

った。

クラスタ 1 には、「エッセイ」のラベルを付与した。出版者や扱われる内容などは定まっておらず、多様なエッセイマンガがこのクラスタに含まれている。

クラスタ 2 には、「手塚治虫全集」のラベルを付与した。このクラスタに含まれるマンガは所蔵館数が下位のマンガも含め全てが講談社の全集に含まれるものであった。一つの全集シリーズであることが、このクラスタを形成する要因と考えられる。

クラスタ 3 には、「ポピュラー」のラベルを付与した。クラスタに含まれるマンガは、人気のあるマンガが主となっている。このクラスタに含まれるマンガは、出版者が限定されていない。また、『今日の猫村さん』のように、少年/少女マンガ・青年/女性マンガに該当しないものも含まれている

クラスタ4には「まんがで読破」のラベルを付与した。このクラスタに含まれるマンガは、所蔵館数下位のものも含め、『資本論』や『神曲』などの古典的な名作をマンガで読む、イースト・プレスのまんがで読破シリーズのものである。

クラスタ 5 には、「シリアス」のラベルを付与した。このクラスタには、原爆や戦争に関するマンガが多く含まれている傾向にある。また、その他、哲学者のニーチェに関するものがある。

クラスタ 6 には、「藤子・F・不二雄全集」のラベルを付与した。このクラスタに含まれるマンガは、下位も含め全て小学館の全集シリーズのものである。

クラスタ7には、「プロジェクト X」のラベルを付与した。このクラスタに含まれるマンガは、2000年代前半に NHK で放映されていたドキュメンタリーテレビ番組(プロジェクト X～挑戦者たち～)に関連した作品群である。このクラスタに含まれるものは、下位も含め全てが、宙出版で発行されたプロジェクトXのシリーズのものであった。

クラスタ 8 には、「タンタンの冒険」のラベルを付与した。このクラスタに含まれるマンガは、下位も含め全てが、福音館書店が発行した、タンタンの冒険シリーズの作品であった。

3.4 クラスタの可視化

八つのクラスタをより深く分析するために、多次元尺度構成法を用いて、クラスタの可視化をおこなった。階層クラスタリングと同様に Simpson 係数に基づく類似度距離を基に、R の `cmdscale` 関数を利用している。

図2は第1次元と第2次元により、図3は第3次元と第4次元によりマンガを配置している。図2において

表1 各クラスタに含まれるマンガタイトル(所蔵図書館数の多い上位5タイトル)

クラスタ1: エッセイ (218タイトル)		クラスタ2: 手塚治虫全集 (106タイトル)	
ペコロスの母に会いに行く	西日本新聞社	ブラック・ジャック 1	講談社
光とともに…: 自閉症児を抱えて 1	秋田書店	鉄腕アトム 1	講談社
わが家の母はピョーキです	サンマーク出版	ブッダ 1	講談社
日本人のちょっとヘンな英語: 爆笑! 英語コミックエッセイ: セイン先生が目撃したおかしな英語	アスコム	火の鳥 1	講談社
戦争論: 新ゴーマニズム宣言special	幻冬舎	アドルフに告ぐ 1	講談社
クラスタ3: ポピュラー (43タイトル)		クラスタ4: まんがで読破 (126タイトル)	
もやしもん 1	講談社	資本論	イースト・プレス
リアル 1	集英社	神曲	イースト・プレス
きょうの猫村さん 1	マガジンハウス	カラマーゾフの兄弟	イースト・プレス
のだめカンタービレ 1	講談社	罪と罰	イースト・プレス
銀の匙 volume 1	小学館	君主論	イースト・プレス
クラスタ5: シリアス (144タイトル)		クラスタ6: 藤子・F・不二雄全集 (39タイトル)	
夕風の街桜の国	双葉社	ドラえもん 1	小学館
ニーチェ: 自由を求めた生涯	筑摩書房	キテレツ大百科 1	小学館
5人の軍隊	金の星社	海の王子 1	小学館
パレスチナ	いそづぶ社	エスパー魔美 1	小学館
ストライク	金の星社	21エモン 1	小学館
クラスタ7: プロジェクトX (19タイトル)		クラスタ8: タンタンの冒険 (17タイトル)	
日米逆転! コンビニを作った素人たち: <セブンイレブン>の流通革命	宙出版	ななつの水晶球	福音館書店
82億食の奇跡: 魔法のラーメン<日清カップヌードル>	宙出版	太陽の神殿	福音館書店
日本初のマイカーてんとう虫町をゆく: <スバル360>家族たちの自動車革命	宙出版	ファラオの葉巻	福音館書店
夢のロータリーエンジン誕生: <ロータリー47士>開発からルマン制覇への挑戦と度胸のトラック便: <クロネコヤマト>の物流革命	宙出版	かけた耳	福音館書店
	宙出版	レッド・ラッカムの宝	福音館書店

は、クラスタ2と4が他のマンガから離れた場所に配置され、クラスタ6もある程度、他のマンガからは距離を置いた場所に配置されていることが読み取れる。また、図4では、クラスタ6,7,8が、他のマンガとは明確に異なる場所に配置されているといえる。

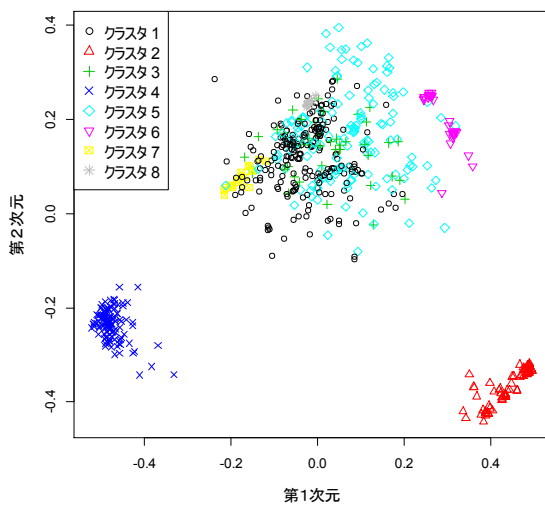


図2 マンガ配置図(第1次元・第2次元)

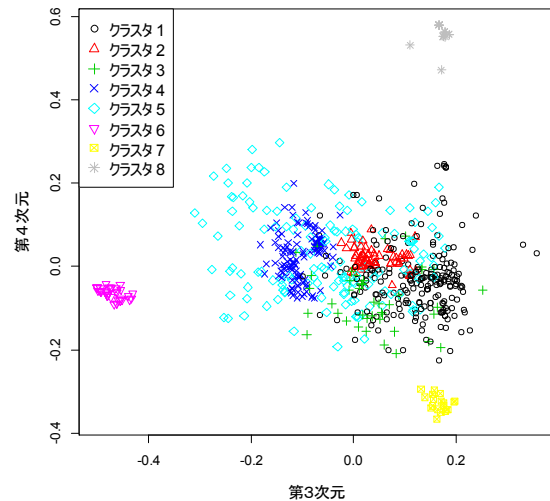


図3 マンガ配置図(第3次元・第4次元)

4. 大学図書館におけるマンガの類型

階層クラスタリングにより作成された八つのクラスタは、大きく二つのタイプに分けることができる。全集やシリーズなどの出版者が作ったマンガのグループがそのままクラスタになっているタイプ(クラスタ2,4,

6, 7, 8)とそれ以外のタイプ(クラスタ1, 3, 5)である。

前者のタイプのクラスタは、図1や図2、図3からもわかるように、それぞれが一つの類型として、大学図書館(員)にとられていることがうかがえる。特に、クラスタ4(まんがで読破)とクラスタ2(手塚治虫全集)は、突出して他のクラスタと分かれており、またクラスタに含まれるマンガタイトル数も多いため明確な類型といえるであろう。また、クラスタ6(藤子・F・不二雄全集)もこれに準じる傾向があるといえる。

後者のタイプのクラスタは、前者のタイプほど明確には分かれていない。図1においては、前者のタイプのクラスタよりも下位においてクラスタが分かれている。また、図2や図3のマンガ配置図においては、クラスタに属するマンガが散らばっており、またクラスタ同士の重なりもみられる。

階層クラスタリングの結果からもわかるように、後者の三つのクラスタは、クラスタ1(エッセイ)・3(ポピュラー)とクラスタ5(シリアス)に分けられる。この三つのクラスタの関係性の解釈の一つは、図3において、クラスタ2(手塚治虫全集)やクラスタ4(まんがで読破)を含めた関係として読み取ることができる。図3では、おおむね、クラスタ2(手塚治虫全集)を境にして、クラスタ1・3とクラスタ4・5に分かれる傾向にある。このことから、エッセイやポピュラーなどの比較的気軽な読み物としてのマンガとまんがで読破シリーズやシリアスなどの学習に関わるようなマンガに分かれ、その中間的な位置づけとして手塚治虫全集がある、と考えることができる。したがって図3からは「気軽な読み物としてのマンガ」や「学習に関わるようなマンガ」が類型として大学図書館(員)に捉えられていることが示唆されたといえる。

5. まとめと今後の課題

本研究では、日本の大学図書館の所蔵データに対してクラスタリングの手法を適用し、マンガの類型化をおこなった。クラスタとしては、八つのクラスタ(エッセイ、手塚治虫全集、ポピュラー、シリアス、藤子・F・不二雄全集、まんがで読破、プロジェクトX、タンタンの冒険)を見出すことができた。

さらにクラスタを分析することにより、マンガの全集やシリーズなどの出版者が作ったマンガのグループがそのまま一つのマンガの類型となる場合があることを明らかにした。また、気軽な読み物としてのマンガや学習に関わるようなマンガといった類型が存在する示唆も得られた。

これらのことから、大学図書館の所蔵情報に対し

て、クラスタリングの手法を適用し、マンガを組織化することの可能性が示されたといえる。たとえば、大学図書館におけるマンガの蔵書評価をおこなうようなツールの開発において、本研究の成果が活用できると考えられる。

今後の課題としては、まず階層クラスタリングにおける下位のクラスタの分析が挙げられる。また、今回のクラスタとそれを所蔵している図書館との関係も検討すべきものである。その他、公共図書館の所蔵データから得られた結果と本稿の結果とを比較し、公共図書館と大学図書館とで、マンガの捉え方にどのように違うのかについて考察することも課題の一つといえる。

注・引用文献

- 1) 原田隆史, 書評中の感性キーワードを用いた小説の分類. 情報知識学会誌. 2005, vol. 15, no. 2, p. 57-62.
- 2) 例えば以下など. Fee, William T.B. Where is the justice... league?: Graphic novel cataloging and classification. *Serials Review*. 2013, vol. 39, no. 1, p. 37-46; *Graphic Novels in DDC: Discussion Paper*. https://www.oclc.org/content/dam/oclc/dewey/discussion/papers/graphic_novels.pdf
- 3) 例えば以下など. Kiryakos, S; Sugimoto, S. A linked data model to aggregate serialized manga from multiple data providers. *ICADL*, 2015. <https://link.springer.com/content/pdf/10.1007/978-3-319-27974-9.pdf#page=134>; 原正一郎, 内藤 求. TOPIC MAPS を利用したマンガ情報の組織化. 研究報告: 人文科学とコンピュータ(CH). 2014, vol. 2014, no. 7, p. 1-8.
- 4) 小山信弥, 吉田倫子, 吉井潤, 上田晶子, 安形輝. “日本の大学図書館におけるマンガの所蔵状況”. 2012年度三田図書館・情報学会研究大会発表論文集. 2012, p. 33-36.
- 5) 安形輝, 杉江典子, 安形麻理, 江藤正己, 大谷康晴, 橋詰秋子. “日本の公立図書館におけるマンガの所蔵状況”. 2016年度三田図書館・情報学会研究大会発表論文集. 2016, p. 25-28.
- 6) Masaki Eto, Teru Agata, Noriko Sugie, Yasuharu Otani, Mari Agata. “Can Japanese manga be automatically classified from public library holdings?” *Proceedings of the 17th ACM/IEEE Joint Conference on Digital Libraries, JCDL 2017, Canada*, p. 287-288.
- 7) NACSIS-CAT 統計情報. <http://www.nii.ac.jp/CAT-ILL/archive/stats/cat/transition.html>