【原著論文】

引用にもとづく雑誌クラスタリング法の開発

天野　晃 *，児玉　閲 **
* 農業生物資源研究所，* 理化学研究所，** 杏林大学，*,** 筑波大学
*amano@brc.riken.jp，**kodamat@ks.kyorin-u.ac.jp

目的：本稿では，雑誌を引用に基づき各学術分野ごとにクラスタリングするための新しい手法を提案する．**対象**：JCR2004 Science Edition CD-ROM 版に収録されている 5964 誌の被引用データより，各誌の被引用比率ベクトルを作成し解析を行なった．**方法**：k-means に基づく，独自に開発したクラスタリングツールを利用した．距離にはコサイン距離（$1-\cos\theta$）を用いてクラスタリングを行なった．**結果**：被引用比率ベクトル空間において各学術分野が形成すると見られるスパイク状の領域が確認された．クラスタリングの結果，安定的に 22 クラスタが同定された．これらは，ほぼ，スパイク状領域に対応しており，それらのクラスタを個々の学術分野に対応付けることができた．著名な総合科学誌は分子生物学関連のクラスタに分類され，これらは被引用傾向において，分子生物学分野の雑誌と同様の特徴を有することが解った．

Clustering Methodology for Journals Based on Citation

Kou AMANO*, Tadashi KODAMA**
*National Institute of Agrobiological Sciences, *RIKEN, **Kyorin University ,
*,**University of Tsukuba
*amano@brc.riken.jp, **kodamat@ks.kyorin-u.ac.jp

Objective: In this paper, we propose a novel method to classify scientific journals into subject fields based on interjournal citation data. **Object**: Citation data of 5964 journals from JCR2004 Science Edition CD-ROM. Cited ratio vectors of the journals were used for analysis. **Method**: Our original clustering programs based on k-means were used. For the clustering, cosine distance（$1-\cos\theta$）was used. **Results**: In the cited-ratio-vector space, the spike-like-regions which appear to correspond to particular scientific fields were observed. And we succeeded to extract each spike-like-region by our clustering method. The number of fields（clusters identified）was 22. It was revealed that the cited feature of prominent multidisciplinary（general）science journals was similar to that of biochemistry journals.

1. はじめに

計量書誌学，特に雑誌評価研究において，雑誌がどのような分野に属するかという問題は重要な要素となる．雑誌評価等によく利用される被引用数やインパクトファクターは，それが属する分野によって大きく異なることが知られているが，複数の分野にかかる評価においては，雑誌が属する分野を適切に区分できるかという問題があり，ひとつの分野を特定する場合にも境界基準の問題が発生する．

ところが，ある雑誌をいずれかの分野に割り当てる方法は限られており，事実上可能な唯一の方法は，いくつかのデータベースにおいて付されている分野を表わすコード（たとえば，Journal Citation Reports（JCR）の Subject Category Code）を利用することである．しかし，この場合においても，分野コードを付与するルールが明らかにされていないことや，本来は利点である，ひとつの雑誌に複数のコードが付与される，あるいは，コー

ドが細分化されていることが，むしろ，問題となる場面がある．

このような問題に対し，本研究では，ある客観的根拠に基づいて雑誌を分類する方法を提案する．

雑誌や文献の分類に対する研究は，従来，雑誌単位よりも記事単位，広範囲にわたる分野よりも特定の分野を対象としたものが多い. 本研究では，自然科学全分野を対象とし，雑誌を主題分野に分類することを考える．

ここで想定する分類の目的は，冒頭に述べたように，雑誌評価その他の計量書誌学研究において，分野別の特徴や傾向の分析に使用することにある．従って，分類項目は 20-30 程度で階層構造も考えないこととする．

論文の分類の場合には，分類ルールのデータ源として，表題，抄録，索引語，本文等に含まれる用語が用いられることが多い．しかし，雑誌の分類では，雑誌タイトルに含まれる用語は限られており，雑誌中の論文に含まれる用語は逆に多様すぎて，適切なサンプルを抽出することが難しい．

これに対し，雑誌間の引用／被引用データは，それらの間の近縁関係についてより有用な情報を与える．ある雑誌がよく引用する（あるいはよく引用される）雑誌は，その雑誌と同じあるいは近縁の主題分野に属するものが多いと考えられる．Garfield の集中則 [1] は，ある主題分野の論文は多数の雑誌に分散するが，そのコアを形成する雑誌はせいぜい 500 誌程度であること，残りの周辺雑誌は他の分野のコア雑誌で有り得ること，を述べたものである．引用関係によって雑誌をクラスタリングすれば，この集中則に近い形で，各クラスタに近縁関係の強いものが集まり，ある分野のコア雑誌群を形成すると期待される．そこで，筆者らは，雑誌間の引用関係を利用したクラスタリングにより，雑誌を分類することを試みた．

このような，多くの分野にわたる大規模な引用データに基づき，分類を試みた研究は希である．Leydesdorff [2, 3, 4, 5] は大規模な雑誌間の関係マップの構築（マップの構築法はたとえば引用／被引用行列に対し主成分分析を施すなどさまざまである）を継続的に行なっており，雑誌の分布を可視化し，雑誌がクラスタを形成することを示したが，雑誌クラスタリングのためのアルゴリズムの提案にまでは至っていない．

Schildt ら [6] は dense network sub-grouping というアルゴリズムを用い，共引用に基づく雑誌のクラスタリングを行なったが，経済・経営分野のみを対象としたもので大規模なものであるとは言えない．

Bassecoulard ら [7] は，引用／被引用行列より，高被引用誌を特定することにより，自然科学分野に対し，効果的に大規模な階層化クラスタリングを行なったが，科学の構造を見ることに主眼が置かれ, そのアウトプットは複雑な階層構造である．

筆者らの研究と同様の観点を持つものには，Glanzel ら [8] が整理した分類があるが，経験に基づく「認知的」なアプローチによるものであり，計量書誌学的な根拠は示されていない．

また, 本研究では自主開発したツールを公開し，多くの研究者に利用可能なものとしたが，計量書誌学の分野では，他の研究者がそのアプリケーションを利用することを想定した実装に関する報告は非常に少く，近年では Schildt ら [6] の報告があるのみであり，しかも，そのアプリケーションはプラットフォームに強く依存するものである．

本研究の特徴は以下の 3 つである．

(1) 自然科学分野全体にわたる大規模な引用関係データに基づく分類法を用いることによって，客観的，定量的なデータに基づく分類結果が期待される．

(2) アルゴリズム（ツール）を自主開発することにより，慣習的な主題分野の考えからは見出されない潜在的分野構造を発見できる可能性がある．また，そうでなくとも，慣習的でない方法により，慣習的に認識される構造を確認できたならば，既存の構造をより強く支持する報告となる．

(3) 本研究の内容と直接関わることではないが，自主開発したツールをオープンソースソフトウエアとして公開することにより，複数のプラットフォームに対応した（あるいは，対応可能な）ツールを多くの研究者に供与できる．

以降の本稿の構成は次のとおりである．2 章において本研究の目的を述べる．3 章において本研究の対象（データの前処理を含む）について述べる．4 章において本研究に用いたデータの特徴について述べる（方法と結果を含む）．5 章においてデータのクラスタリングについて述べる（方法と結果を含む）．6 章において以上をまとめた考察を述べる．

2. 目的

本研究の目的は，1. 雑誌間の引用／被引用の特徴を明らかにし，2. これに基づくクラスタリング方法を提案し，3. これを用いた実験により提案方法の妥当性を確認する，ことにある．

筆者らは，当クラスタリング方法が，他の方法や前述したデータベースによって付与されるカテゴリによるものと比べ，最善の方法であると主張するものではない．ここでの提案が，計量書誌学研究に対し多角的解析／評価のためのひとつの新しい視点を提供しその一助となることを意図している．

3. 対象

Thomson Scientific 社（現 Thomson Reuters Scientific 社）が提供する，Journal Citation Reports（JCR）2004 Science Edition CD-ROM 版（以降，これを単に JCR と呼ぶ）に収録されている雑誌について，これらの被引用ベクトルを作成し，これを分類の対象とする．JCR の Cited Journal Data は，各雑誌が当該年に受けた被引用回数を，引用元の雑誌別，被引用文献の発行年別に示している．まず，これらから引用元誌（以降，これを単に引用誌と呼ぶ）が特定されていないデータを除いた（これらのデータは，ある被引用誌を 1 回しか引用しなかった引用誌についてのものであり，Thomson Scientific 社はこれらを一括して All Others として集計している）．残ったデータから，発行年が 2000-2004 年である論文に対して少くとも 1 回の引用がある被引用誌－引用誌対を抽出すると，そのなかに 5964 誌が存在した．これらのデータを用い，5964×5964 の被引用／引用行列を作成した．この行列から得られる 5964 の被引用誌ごとの被引用ベクトルを，雑誌クラスタリングのためのデータとする．被引用ベクトルの各要素は，当該被引用誌の 2000-2004 年の論文が，各引用誌から 2004 年に受けた被引用数である．

しかし，非常に多くの被引用数を持つ一部の雑誌や，他誌からの引用に比べ極端に自誌引用の多い雑誌が，分野間の被引用の特徴の差を上まわる影響を及ぼすと考えられるため，1 章に述べたような特徴をうまく引き出すには，これらを緩和する必要がある．そこで，以下のようなデータの前処理を行なった．

前処理 1. 自誌引用の影響の緩和

まず，それぞれの被引用誌に対し最大の被引用数を持つ引用誌が自誌であるとき，その被引用数を次に大きい被引用数と同じ値に変更した．ただし，次に大きい被引用数が 0 でないときに限る．

前処理 2. 高被引用誌の影響の緩和

次に，それぞれの被引用誌における各引用誌からの被引用数を，その被引用誌の全被引用数で除したのち 1000 を乗する．つまり，被引用比率のみを考慮する．

以降，この前処理を行なった後のベクトルを被引用比率ベクトルと呼ぶ．

4. データの特徴の解析

4.1 作業仮説

被引用比率ベクトル空間の中で，各雑誌は，ベクトルの要素和が 1000 となる条件を満たす超平面内に存在する．1 章で述べたように，近縁関係によって雑誌がいくつかの主題分野にクラスタ化されるとするならば，この超平面内には，同じ分野に属する雑誌が集中するいくつかの領域が形成されると考えられる．すなわち，各領域がある分野に対応する．この考えに立ち，以下の仮説を設定する．

仮説 1. 被引用比率ベクトル空間中の超平面（ベクトルの要素和 =1000 の条件を満たす）内に，雑誌が集中したいくつかの領域ができる．各領域に集まる雑誌は同じ分野に属する．

仮説 2. 異なる領域は異なる分野に対応するので，2 つの領域の中心（その領域に属する雑誌の座標の重心）が，被引用比率ベクトル空間の原点を挟む角はある程度の大きさを持つ．

仮説 3. 多数の分野から引用される雑誌は，被引用比率ベクトルの各要素値（各雑誌からの被引用数）が比較的均等な値になるため，原点からの距離が近い．このような雑誌は，被引用数の多い雑誌が中心である．

4.2 方法

上記の仮説を検討するため次の調査を行なった．なお，ここで検討できるものは仮説 2 および仮説 3 のみである．仮説 1 の検討は次章で行う．

調査 1. 被引用比率ベクトル空間内の雑誌の分布（プロット）

調査 2. 被引用比率ベクトル空間における原点から各雑誌のベクトル（座標）へのユークリッド距離

調査 1 を行うにあたって，超平面空間である被引用比率ベクトル空間内の雑誌の分布をプロットする場合，超平面空間を（現実的に表現可能な）せいぜい 3 次元の空間へ投影する必要がある．この場合，分布の特徴を最もよく表すには，被引用比率ベクトルに対し主成分分析を行い（この主成分分析によって得られるベクトルを被引用比率主成分ベクトルと呼ぶ），その上位軸の座標値で雑誌をプロットするのがよい．主成分分析に分散共分散法を用いれば，主成分分析前後におけるベクトル空間内の雑誌の相対的な位置関係は変化しない．調査 1 では，被引用比率主成分ベクトル空間の上位 1-3 軸および 4-6 軸の座標値を用いたプロットを行い，雑誌の分布の状態を確認する．

ところで，被引用比率ベクトル空間における原点は，全ての被引用比率の値が 0 となるベクトル（ゼロベクトル）であるが，被引用比率主成分ベクトル空間における原点は，もとの被引用比率ベクトルの重心（平均）であり，一致していない．そこで，本文では混乱を避けるため，被引用比率ベクトル空間におけるゼロベクトルおよび被引用比率主成分ベクトル空間においてこれと対応する点を，「原点」，被引用比率主成分ベクトル空間における原点およびもとの被引用比率ベクトルの重心を，「重心」と呼ぶ．

4.3 結果

被引用比率ベクトルに対し主成分分析を行ない，上位軸についてのプロットを行なったものを図 1 に示す．同図より，雑誌が集中する領域は，被引用比率主成分ベクトル空間において，放射状に伸びるスパイクとして確認できる．スパイクの一端が集中しているように見える点が原点である．また，それぞれのスパイクは明確に分離している（原点から伸びる向きが明確に異なる）こと

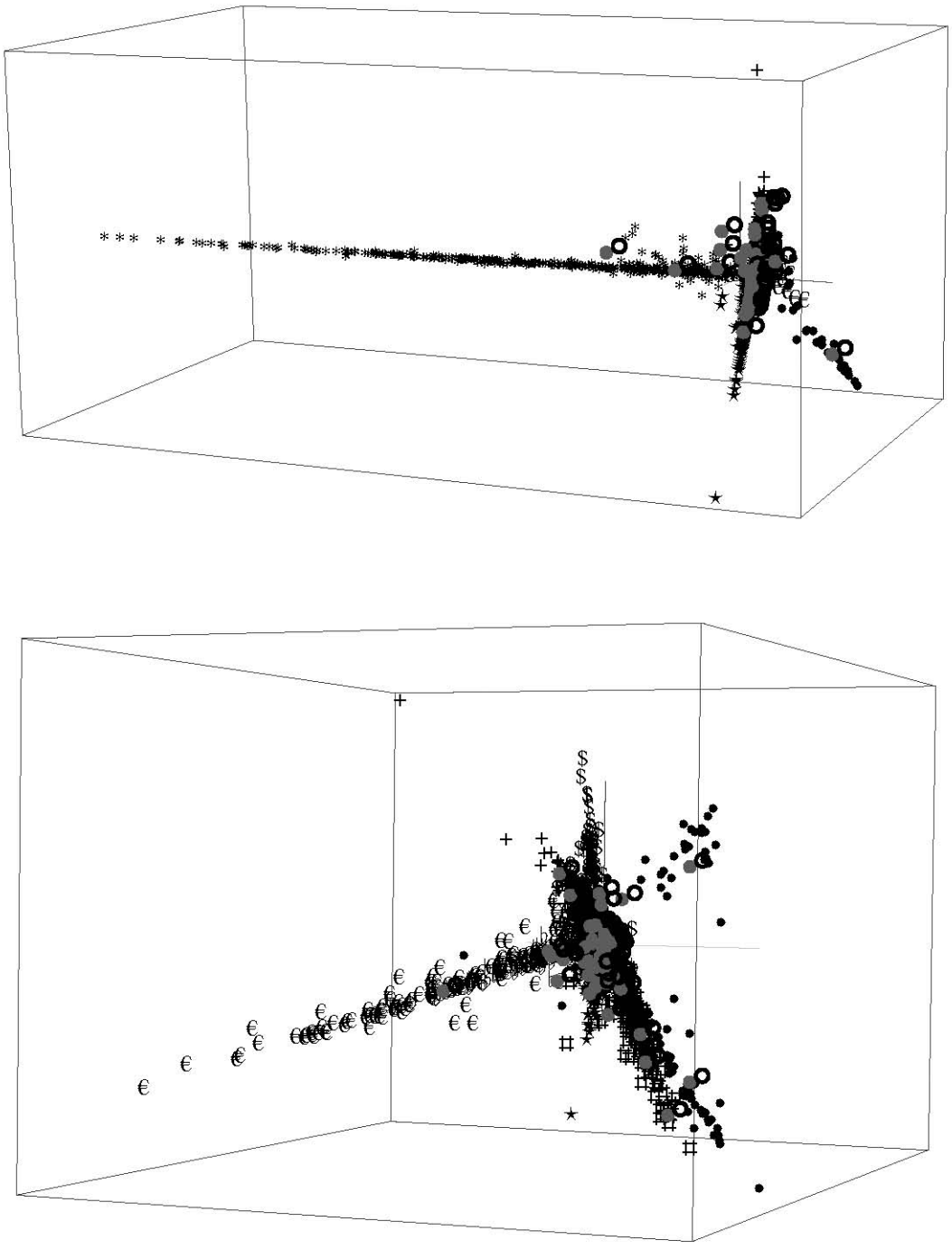

図 1：被引用比率主成分ベクトル空間における雑誌の分布（3D 表示）*

*上段は第 1，2，3 軸，下段は第 4，5，6 軸によるプロット．各マークは雑誌を表わし，同じクラスタに属する雑誌には同じマークが付される．ただし，被引用数合計において上位 20 位まで，あるいは，JCR カテゴリが総合科学である雑誌については，グレイのドットで表わされる．空間全体に学術分野ごとに形成されると見られる明らかなスパイクが確認できる．

が確認できる．このようなスパイクはさらに下位の主成分軸においても確認された. このことより，仮説 2 が支持される．

被引用比率ベクトル空間における原点から各雑誌への距離のヒストグラムを図 2 に示す．雑誌の集中は原点からの距離にして 200 付近に見られ，前後 100 の距離（100-300）の間に全雑誌の約 73%が存在する．

また，原点付近に位置する雑誌と原点から遠くに位置する雑誌との被引用数の傾向を見るため，図 2 に各ヒストグラムカテゴリごとの総被引用数を示す．同図は，雑誌が最も集中して存在する領域（原点からの距離にして 150-200）よりも，総被引用数が最も集中する領域（原点からの距離にして 100-150）の方がより原点に近いことを示している．このことより，仮説 3 は支持される．

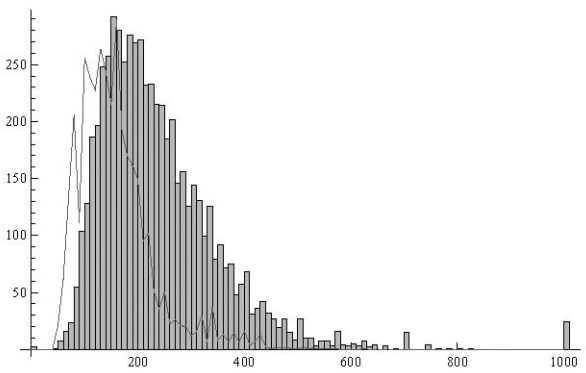

図 2：被引用比率ベクトル空間における原点からの距離のヒストグラム*

*実線は各ヒストグラムカテゴリに属する雑誌の総被引用数．横軸は被引用比率ベクトル空間における原点から各雑誌へのユークリッド距離．縦軸はヒストグラムの頻度および総被引用数（軸表示×2000）.

5. クラスタリング

5.1 作業仮説

すでに述べたように，被引用比率ベクトル空間におけるこれらのスパイクがなんらかの学術分野を表わすと考えることは妥当であり，従って，各学術分野を分離するためには，各スパイクを分離するクラスタリングを行なえばよい．

このような，原点から放射状に伸びる領域を分離するクラスタリングを行なうには，クラスタ間の距離にコサイン距離（$1-\cos\theta$）を用いるのが適している．また，学習型非階層分類法を用い，初期段階では各スパイクの中腹部付近にクラスタを配置し，続く学習過程においては極端に近接するクラスタを統合するクラスタリングを行なうことにより，少ない計算量で妥当な（ロバストな）結果を得られる．

5.2 方法

初期クラスタの配置についてはひとつ問題がある．作業仮説にあるようにスパイクの中腹に初期クラスタを配置するのに，まずスパイクを検出しなければならないが，そもそもスパイクを検出するためにクラスタリングするわけであるから，これでは矛盾してしまうというものである. そこで，筆者らは，被引用比率ベクトルに対し主成分分析（この場合も分散共分散法を用いるのが適している）を行なえば，各軸が各スパイク領域と重なるかおおむね近くに配置されると考え，上位の主成分軸上に初期クラスタを配置することにした．仮にいくつかの初期クラスタがスパイク領域からはずれた場合でも，学習において遠距離にあるクラスタは利用されないようなアルゴリズムを採用すれば問題は無い．

以上の条件を満たすため，筆者らは Self-organizing clustering（SOC）[9, 10, 11] を用いクラスタの初期配置と学習を行なった．このプログラムは学習型 k-means を基盤とするが，その弱点である初期値依存性を回避し得るアルゴリズム，および，動的にクラスタの統合／生成を行なうそれが組みこまれている．しかしながら，SOC には，コサイン距離を用いるクラスタリングなど，提案手法に必要となるいくつかの機能が用意されていなかったため，筆者らはこれらの機能の追加を行なった [12]．これらのアルゴリズムを使用した場合にお

いても学習型 k-means のオーダー（$O(cnml)$：このとき，c：クラスタ数，n：サンプル数，m：ベクトルの大きさ，l：ループ回数）を維持できる．

以上を勘案し，以下の手順でクラスタリングを行なった．

手順 1. 被引用比率ベクトルに対し主成分分析を行なう．

手順 2.「被引用比率主成分ベクトル空間におけるゼロベクトルと雑誌ベクトルとの関係」と「被引用比率ベクトル空間におけるゼロベクトルと雑誌ベクトルとの関係」が同等になるよう，被引用比率主成分ベクトル全体をシフトさせる．
この操作を行わないと，「雑誌間の重心を挟む角（コサイン距離）」を得ることになり，「雑誌間の原点を挟む角」が得られなくなる．クラスタリングで期待するのは後者である．

手順 3. シフト後の上位 32 位までの主成分軸上における雑誌の平均位置に初期クラスタを配置する．

手順 4. コサイン距離を用い，閾値（コサイン距離にして 0.75）よりも近接するクラスタを統合しながら，十分にアニーリングを行なう．

なお，被引用ベクトルがゼロベクトルとなる雑誌が 2 誌存在したが，これらを除いた 5962 誌をクラスタリングの対象とした．

クラスタが適切なサイズを持ちかつ互いに明確に識別されるかを検証するため，以下の調査を行なった．

調査 1. クラスタをメンバ数順に並べたときのランクとメンバ数の関係

調査 2. 各クラスタにおける JCR カテゴリ間分布から計算されるクラスタ間相関係数

調査 3. JCR カテゴリにおけるクラスタ間分布から計算される各 JCR カテゴリの Gini 係数

調査 4. 各クラスタ重心間のコサイン距離

調査 2 と調査 3 の手順は次のようになる．対象誌に対しクラスタと JCR カテゴリをひとつずつ付与する．クラスタは 1 誌につきひとつが付与されるが，JCR カテゴリは 1 誌につき複数付与されるものがあるので，これらの雑誌については，付与されている JCR カテゴリの数と同じ数になるよう同名の雑誌を増やし，それぞれの JCR カテゴリとクラスタの対を作る．これに対しクラスタと JCR カテゴリとで雑誌数をクロス集計する．クロス集計表を，JCR カテゴリ分布を示す各クラスタのベクトルと考えると，クラスタ間の相関係数を得ることができる．逆に，クラスタ間分布を示す各 JCR カテゴリのベクトルと考えると，それぞれの JCR カテゴリのクラスタ集中性を示す Gini 係数を得ることができる．

クラスタ間の相関が十分に低ければ，それぞれのクラスタが異なる分野を表わすことが示唆される．また，各 JCR カテゴリの Gini 係数が大きい（1 に近い）ほど，そのカテゴリの雑誌がひとつないし少数のクラスタに集中していることを示す．

5.3 結果

最終的に 22 クラスタが得られた．初期クラスタの配置を変化させた場合（上位 31-28 軸）にも，同様に 22-23 クラスタが安定して得られることを確認した．

各クラスタがおおよそどのような分野と対応するかを確認したものを表 1 に示す．対応関係は，各クラスタに含まれる雑誌の，誌名とそれに与えられる JCR カテゴリとから判断した．また，そのクラスタ内での最多被引用誌が，判断される分

表 1：各クラスタと対応する学術分野

クラスタ	メンバ数	おおよそ該当する学術分野	クラスタ内の最多被引用誌	原点からクラスタ重心までの距離	一誌当りの被引用数
M	216	数学	*Chaos Solitons Fractals*	78	177
Q	78	素粒子・核・天体物理	*Astrophys J*	172	3473
G-1	308	気象・環境	*J Geophys Res*	65	657
G-2	165	地球科学	*Earth Planet Sci Lett*	81	465
C-1	315	化学	*J Am Chem Soc*	76	2283
C-2	248	物性物理	*Phys Rev Lett*	83	2203
Ca	130	分析化学	*Anal Chem*	99	986
Cp	163	応用化学	*Macromolecules*	90	761
B-1	427	動植物学	*Plant Physiol*	54	590
B-2	243	農学・食品	*Appl Environ Microbiol*	67	607
B-3	148	水産・海洋	*Mar Ecol-Prog Ser*	84	464
Mo	751	生化学・生理学	*J Biol Chem*	53	3208
Me-1	416	臨床医学 1	*Gastroenterology*	43	878
Me-2	519	臨床医学 2	*New Engl J Med*	41	1344
Me-3	374	心理・神経科学	*J Neurosci*	54	1339
Me-4	277	獣医学・感染症	*J Clin Microbiol*	58	810
E-1	209	金属・セラミックス	*J Electrochem Soc*	79	514
E-2	187	機械・土木建築	*J Acoust Soc Am*	70	270
E-3	173	化学工学・エネルギー工学	*Ind Eng Chem Res*	71	287
I-1	433	電子情報工学	*Lect Notes Compt Sci*	225	252
I-2	178	数理工学	*IEEE Trans Microw Theory Tech*	66	169
I-3	4	信頼性工学	*Reliab Eng Syst Saf*	346	115

野と矛盾しないことを確認した．

クラスタをメンバ数順に並べたときの，メンバ数とランクの関係を，図 3 に示す．最小クラスタのメンバは 4 誌，最大クラスタのメンバは 751 誌であったが，この 2 クラスタを除くとメンバ数は 78-519 の範囲にあり，分布は比較的均等である．このときの，ランク r とメンバ数 $N(r)$ の関係を Zipf の法則 $N(r)=Cr^{-a}$ に当てはめると，指数 a は約 0.5 で，通常の Zipf 型分布で見られる $a \approx 1$ よりかなり小さい．また，この分布の Gini 係数は 0.33 である．これらのことから分布が少数のクラスタに集中しておらず，比較的均等であることが判る．

231（$22\times21\div2$）組のクラスタ間の相関は，0.5 以上となるものは存在せず，0.5 未満 0.4 以上のも

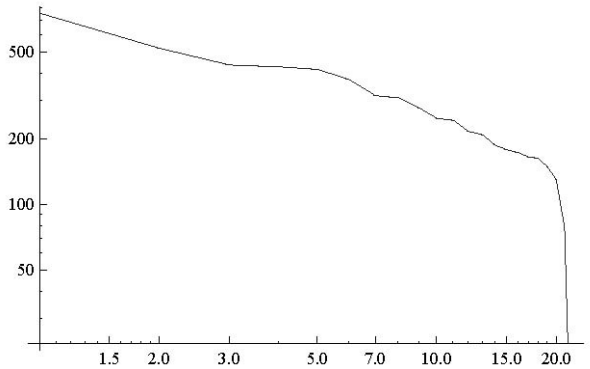

図 3：クラスタメンバ数とそのランク*

*横軸がランク，縦軸がメンバ数による両対数プロット．多くのクラスタのメンバ数は，100 から 500 の間にあり，一様分布に近い．

のが 1 組（クラスタ I-1 と I-3），0.4 未満 0.3 以上のものが 3 組，0.3 未満 0.2 以上のものが 6 組，0.2 未満 0.1 以上のものが 11 組，0.1 未満 −0.1 以上のものが 208 組，−0.1 未満 −0.2 以上のものが 2 組であった（以上, 全 231 組）．有意な相関（$p<0.05$）はひとつもなく，各クラスタが異分野に分類されていることが示唆された．

JCR カテゴリの Gini 係数は，0.95 以上のものが 81, 0.95 未満 0.9 以上のものが 51, 0.9 未満 0.85 以上のものが 23，0.85 未満のものが 15（以上，全 170 カテゴリ）で，ほとんどの JCR カテゴリにおいて，その属する雑誌はごく少数のクラスタに集中している．Gini 係数が低いカテゴリは，総合科学（0.52）, 学際応用物理（0.68）, 生物学（0.71），総合工学（0.73）等，広い分野にまたがる分野である．

231 組のクラスタ重心間のコサイン距離は，0.7 未満となるものは存在せず，0.7 以上 0.8 未満のものが 8 組，0.8 以上 0.9 未満のものが 26 組，0.9 以上のものが 197 組であり（以上，全 231 組），各クラスタ中心が十分に分離されていることが確認された．

また，前掲の図 1 においては，各雑誌は属するクラスタごとにマークされている．同図からは，おおむね，ひとつのスパイクがひとつのクラスタに属していることが判る．

以上より，前掲 4 章仮説 1 が支持される．

6. 考察

筆者らの提案する学術分野の特定方法（雑誌の被引用数に基づく k-means をベースとしたクラスタリング）は以下の理由から適切であったと言える．

（1）クラスタのサイズが比較的均等であった．

（2）各 JCR カテゴリ内の雑誌の分布を見たとき，雑誌が 1 つないし少数のクラスタに集中し，また，そうでないものは総合科学等，容易に説明ができる分野である．

（3）有意なクラスタ間の相関はなく，クラスタがよく分離されていた．

（4）各クラスタ内の JCR カテゴリ分布から，それぞれのクラスタの分野が無理なく定義できた．

（5）雑誌を個別に見た際に，不適切と思われるクラスタへ所属しているものはほとんど存在しなかった．

また，図 1 において，JCR カテゴリの総合科学分野に属する雑誌と，（被引用ベクトルより）自誌引用を含めた被引用数の合計の順位で上位 20 位までに含まれる雑誌をマークした．これを見ると，これらのうちの全てが原点付近に集中するようなことはなく，いずれかのスパイクの中腹付近にかけての分布が確認できる．このことから，多くの総合科学誌や高被引用誌が，いずれかの学術分野と同じ被引用傾向を持つことが判る．

以下，これらのクラスタについて，概観や特徴を述べる．記述は，そのテーマから見て，近いと思われる分野ごとにまとめた．また，これらは，実際にクラスタ重心間の距離が，比較的近いものである．

クラスタ M（数学）：数学または数理物理学を主なテーマとするクラスタ．

クラスタ Q（素粒子・核・天体物理）：いわゆる天文学と量子力学をテーマとするクラスタ．現在の天文学の特徴を良く表わしていると言える．

クラスタ G-1，G-2（環境・地球科学）：クラスタ G-1 には気象・環境学が，G-2 には地球科学が分類された．クラスタ G-1 には土壌学，水資源学を, G-2 には地質, 鉱物, 古生物学の分野を含む．

クラスタ C-1，C-2（化学）：クラスタ C-1 に化学一般，C-2 に物性物理学が分類された．クラスタ C-1 に化学物理を, クラスタ C-2 に光学を含む．

クラスタ Ca（分光学・分析化学）：分析化学を

主なテーマとするクラスタ．ここに法医学が含まれることは興味深い．

クラスタ Cp（応用化学）：高分子化学と応用化学を主なテーマとするクラスタ．

クラスタ B-1，B-2，B-3（農学・生物学）：クラスタ B-1 に動植物学が，B-2 に農学・食品学が，B-3 に水産・海洋学が分類された．クラスタ B-1 には生態学，森林学，農業経済の分野を，B-2 には園芸を含む．

クラスタ Mo（分子生物学）：最大のクラスタ（メンバ数 751）であり，生化学，生理学を主なテーマとする．生物物理，遺伝，薬学，病理，免疫，内分泌，血液，腫瘍，泌尿器の分野を含む．Nature および姉妹誌，Science，Proceedings of the National Academy of Sciences of the United States of America といった，総合科学誌の中でも，インパクトファクターが高く大きな影響を持つ雑誌はここに含まれる．これらの雑誌が，被引用において同様の傾向を持つということは，そこに含まれる記事の内容においても同様の傾向があると考えられる．つまり，総合科学誌の中でも特に影響の大きなものは生化学・生理学に関わる記事を多く含むことが示唆される．

クラスタ Me-1，Me-2，Me-3，Me-4（医学）：クラスタ Me-1 と Me-2 に臨床医学が，Me-3 に心理・神経科学が，Me-4 に獣医学・感染症学が分類された．

クラスタ Me-1 には消化器，外科，皮膚科，耳鼻咽喉科，リウマチ，リハビリ，放射線医学，歯学，生体材料の分野を，Me-2 には呼吸器，循環器，小児科，産婦人科，麻酔，救急医療，看護，ヘルスケア，公衆衛生の分野を，Me-3 には行動科学を，Me-4 にはウイルス，寄生虫，眼科の分野を含む．クラスタ Me-1 はどちらかといえば外科系の，クラスタ Me-2 はどちらかといえば内科系の医学であるように見える．

クラスタ E-1，E-2，E-3（工学）：クラスタ E-1 に金属・セラミックス関連が，E-2 に機械・土木建築学が，E-3 に化学工学・エネルギー工学が分類された．

クラスタ E-1 には鉱山学を，E-2 には音響学，航空宇宙分野を，E-3 には熱力学，科学哲学の分野を含む．科学哲学は熱力学やエネルギーとの関係が強いという印象があるが，引用関係からもこれが示されたことは興味深い．

クラスタ I-1，I-2，I-3（情報学・計算機科学）：クラスタ I-1 に電子情報工学，I-2 に数理工学，I-3 に信頼性工学が分類された．クラスタ I-1 は全 22 クラスタの中で最も顕著なスパイクを示した．クラスタ I-3 は雑誌数がわずか 4 という最小のクラスタであるが，明らかに信頼性工学の雑誌から成っており，これが独立したクラスタを構成したことは興味深い．

クラスタ I-1 にはロボティクス，通信工学を，I-2 には確率･統計，電力･電送工学，製造，輸送，OR の分野を含む．いずれのクラスタも一誌あたりの被引用数は低い．

この調査より明らかになった最も興味深いことは，一般的に著名な総合科学誌の多くの被引用の傾向は，分子生物学関連の雑誌のそれと同様であるということであろう．このことは，現在の科学における分子生物学への関心の高さを示している．このような傾向は総合科学誌に含まれる記事を見れば推測されることであるが，当結果はこのことが引用分析からも確認できたひとつの例となる．なお，前述したように JCR カテゴリの総合科学の Gini 係数は低く（0.52），全体として見ると，この分野の雑誌はさまざまなクラスタに分離している．

最後に，筆者らの提案は雑誌集合を各学術分野にクラスタリングするためのひとつの方法であり，各分野におけるコアジャーナルや重要な雑誌の決定にまで及ぶものではなく，そのためには，インパクトファクター等，さらに多くの指標や，本稿では述べなかった異なる観点が必要であるということを述べておく．

謝辞

筑波大学図書館情報メディア研究科小野寺夏生教授には研究の全般において非常に多くの指導と助言を頂きました．深く感謝いたします．筑波大学図書館情報メディア研究科岩澤まり子教授，深海薫教授，眞榮城哲也准教授，石川大介氏には研究の全般において多くの助言を頂きました．深く感謝いたします．プログラムの開発／テストには，農林水産省農林水産技術会議科学技術計算システムを利用させて頂きました．深く感謝いたします．

注・文献

[1] Garfield, E. Citation indexing: its theory and application in science, technology and humanities. New York, Wiley, 1979, p. 23.

[2] Leydesdorff, L. Various methods for the mapping of science. Scientometrics. Vol. 11, Nos 5-6, 1987, p. 295-324.

[3] Leydesdorff, L. Clusters and maps of science journals based on bi-connected graphs in Journal Citation Reports. Journal of Documentation. Vol. 60, No. 4, 2004, p. 371-427.

[4] Leydesdorff, L. Indicators of structural change in the dynamics of science: Entropy statisitics of the SCI Journal Citation Reports. Scientometrics. Vol. 53, No. 1, 2002, p. 131-159.

[5] Leydesdorff L. Can scientific journals be classified in terms of aggregated journal-journal citation relations using the Journal Citation Reports? Journal of the American Society for Information Science and Technology. Vol. 57, No. 5, 2006, p. 601-613.

[6] Shildt, H. A. and Mattsson, J. T. A dense network sub-grouping algorithm for co-citation analysis and its implementation in the software tool Sitkis. Scientometrics. Vol.67, No. 1, 2006, p. 143-163.

[7] Bassecoulard, E. and Zitt, M. Indicators in a research institute: A multi-level classification of scientific journals. Scientometrics. Vol. 44, No. 3, 1999, p. 323-345.

[8] Glanzel, W. and Schubert, A. A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. Scientometrics. Vol. 56, No.3, 2003, p. 357-367.

[9] 天野晃．自己組織化によるDNAシークエンスの分類．情報メディア学会第5回研究会発表資料．2003，p. 5-8.

[10] Amano, K., Nakamura, H. and Ichikawa, H. Self-organizing clustering: A novel non-hierarchical method for clustering large amount of DNA sequences. Genome Informatics. Vol. 14, 2003, p. 575-576.

[11] Amano, K., Ichikawa, H., Nakamura, H., Numa, H., Fukami-Kobayashi, K., Nagamura, Y. and Onodera, N. Self-organizing clustering: Non-hierarchical clustering for large scale DNA sequence data. IPSJ Digital Courier. Vol. 3, 2007, p. 193-197.

[12] 配布については amano@brc.riken.jp まで連絡願いたい．

（2008年7月3日　受付）
（2008年10月22日　採録）
（2008年12月26日　出版）