

「個」の同定と認証—分類の視点から—

小沢 一雅 (大阪電気通信大学)

人・もの・情報における「個」の同定と認証について、集合の分類という視点から考察する。まず、同定と認証という2つの機能について明確な定義を与える。つぎに、これらを実行する場としての作業空間のあり方を考察し、理想的な作業空間を構成する手順を考える。例題集合として喜多川歌麿・葛飾北斎・安藤広重の美人画における顔のコレクションを選び、「個」の同定と認証について実験的に考察する。

Identification and Authentication of An Individual Belonging to A Set

Kazumasa Ozawa

Osaka Electro-Communication University,
Neyagawa, Osaka 572-8530, Japan.

Email: ozawa@ozlab.osakac.ac.jp

This paper defines identification and authentication of an individual in a set of objects in the light of taxonomy. Namely identification is to classify a given individual into a known subset. Authentication is nothing but verification or proof of the hypothesis that a given individual belongs to a known subset. Important is that a different mathematical space would be required for doing each of them. Finally identification and authentication of an individual are discussed by employing an example set of female faces extracted from *Ukiyoe* wood-printed pictures originally drawn by the three well-known painters *Utamaro*, *Hokusai* and *Hiroshige*.

1. はじめに

人・もの・情報における「個」をいかに同定し、認証するかが情報社会における重要な課題になると予想される。情報社会では、実体としての「個」をさまざまな形式で電子的に記述した記号（電子化記号）によって、その「個」を代替させる状況が一般化するのとは必然的とみられる。たとえば、個人認証やデジタルアーカイブなど昨今注目されている技術動向もこれに関連する。本来、実体としての「個」のもつ情報量は無限大であって、似たものの集団の中にあっても必ず他者を明確に区別する情報（固有情報）を含んでいると考えられる。「個」の同定と認証を電子化記号のレベルで行おうとする場合には、まず電子化記号がそれをやるにふさわしい固有情報を含んだ形式になっていなければならない。

一方、実体のレベルであれ電子化記号のレベルであれ、「個」の同定と認証を行うにあたっての基本原理に関していえば、一貫した普遍性があると考えられる。すなわち、同定と認証に必要な情報をいかに効果的に抽出するか（できるか）という一点につきる。数理的

な用語でいい換えれば、必要な情報を効果的に抽出する空間（座標系）の選択の問題に帰着するといえよう。

人文科学の世界においても、本質的に「個」の同定と認証に関連する具体的な事例はじつに多い。たとえば、文学作品の作者の同定、古文書の筆跡鑑定、あるいは各種の真贋鑑定などがあげられよう。問題は、これらの事例がそれぞれ独立した無関係な事例として受けとめられ、たがいに関連づけられることなく個別の経験的な視点にもとづいてとりあつかわれてきたことである。

本稿は、「個」の同定と認証を人文科学の世界におけるこうした個別的な事例を総合するテーマとして位置づけつつ、まずは分類の視点から基礎的な考察を述べる。そもそも「個」とは他者を区別することによって生まれる概念であり、それにはものの分類という視点が不可欠だからである。

2. 集合と「個」

人・もの・情報について、さまざまな集合がある。集合とは個体の集まりであって、個々の個体は簡潔に元とよばれる。人の集合についていえば、1つの元とは1人の個人をさすことになる。万葉集は奈良時代に編纂されたの歌の集合（歌集）であって、1つの元は1つの歌をさす。一方、イヌやネコといった動物名は、哺乳動物の全個体の集合を種によって分類した結果生まれる1つの部分集合に対応している。万葉集から柿本人麻呂の歌を1つとりだすと1つの元であるが、彼の歌全体をとりだすと1つの部分集合をなす。この場合、いずれもが「個」としてとりあつかわれる可能性をもっている。分類の視点からいえば、こうした元と部分集合の並立状態は好ましくない。そこで、柿本人麻呂の歌を1つとりだす場合、これを1つの元からなる部分集合（シングルトン）をとりだすこととみなす。

「個」を考える具体的な場面では、背景に必ず何らかの集合があり、その部分集合が重要な役割を演ずる。こうした集合とのかかわり方の実例を総合して、本稿では「個」をつぎのように定義する。すなわち、「個」とは、集合から選ばれた1つの部分集合である。

集合における1つの部分集合を「個」としたとき、「個」以外のすべて（補集合）を他者とよぶ。すなわち、「個」の成立は同時にそれとは区別される他者の発生をひきおこす。ここでいう区別とは、集合の分類（分割）にほかならず、集合を「個」と他者という2つの部分集合に分割することである。

2. 1 同定と認証

一般に、正体不明のものが与えられたとき、それが何であるかを決定する操作を同定という。集合との関連でいうと、与えられた個体がどの部分集合に属するのかを決定する操作である。つまり、目の前の動物がイヌか、ネコか、あるいはその他のどの種に属する個体なのかを決めるのが同定である。

一方、認証とは、与えられた個体が特定の部分集合に属する元であるとする仮説があったとき、それを検証する操作である。つまり、目の前にいる動物がイヌだというけれど、それが本当なのかどうかを確かめる操作を認証とみなす。

上のように規定した同定と認証のちがいは、以下のようにそれぞれを設問と回答の組のかたちでとらえると明瞭になる。

(1) 同定

【設問】 What is it? 【回答】 It is A . (A は部分集合, 厳密には It belongs to A)

(2) 認証

【設問】 Is it A ? 【回答】 Yes (or No).

ここで、分類の視点から「個」の同定と認証について考えるため、まず典型的な「個」の同定問題をとりあげる。集合 M が n 個の部分集合に分割され、部分集合のそれぞれが異なった「個」を表徴している場合である。すなわち、あらかじめ複数の「個」の存在を想定する場合であって、不明個体 x がどの部分集合に属するかを判定することによって x が何であるか(どの「個」であるか)が決定される。一方、「個」の認証とは、ただ1つの「個」と他者と

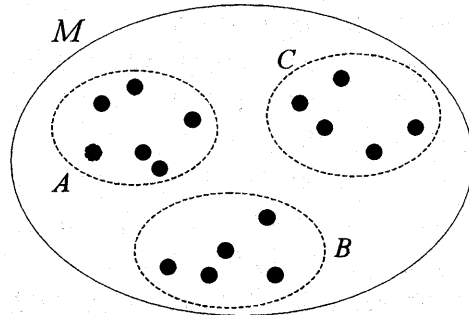


図1 集合と部分集合

の類別問題である点で同定とは対照的なちがいがあ。すなわち、おなじ集合に関して、同定とは n 個の場合わけ問題 (n 分類) であり、認証とは2個の場合わけ問題 (2 分類) である点にちがいがあ。図1に、集合 M が3つの部分集合 A, B, C に分割されている例を示している。不明個体 x の同定とは、 x が3つの部分集合 A, B, C のいずれに属するかの類別問題であるのに対して、認証とは $x \in A$ という仮説の検証であって、 x が「個」 A と他者 $B \cup C$ という2つの部分集合のいずれに属するかという類別問題になる。

2. 2 作業空間

「個」の同定と認証のちがいは、それを実行するための作業空間の差異にもつながっていく。ここでいう作業空間とは、集合の元たちを数量的に比較するための場としての座標系(座標軸の組)をさす。元たちを観察する視点として1つの属性を導入すれば、その観測値(属性値)を表現する1つの座標軸が設定できる。たとえば、人の集合を考えるにあたって、身長、体重、および足の長さといった3つの属性を導入すれば、3つの座標軸からなる作業空間ができあがる。問題は、こうした作業空間が同定と認証のいずれを行うのかによってちがってくる点である。

図1の例で考えてみる。不明個体 x の同定では、 x が3つの「個」 A, B, C のいずれにも該当する可能性をもっているため、作業空間には3つの「個」の固有情報がすべてまんべんなく反映される座標軸を設定する必要がある。一方、 $x \in A$ を認証するためには、作業空間として A の固有情報がきちんと表現できる座標軸を設定することが重要であるが、ここでは、 B と C の類別を考える必要はない。無関係な情報は認証にとってたんに不要というだけではなく、むしろ阻害要因になるため積極的に排除していく必要がある。

「個」の同定と認証において作業空間が異なる点を指摘した。一方、それぞれを実行する処理手順という側面で見れば、共通する部分もある。一方は n 分類であり、他方は2分類であって、両者とも分類操作にほかならないからである。数理的な分類において距離尺度が重要な役割を演ずるのは周知のとおりであるが、同定と認証においても当然適当な距離

尺度の導入が必要になる。一般には、作業空間における2つの元の距離は、それぞれの座標値を総合するかたちで定義される。

いま、集合 M が n 個の「個」の候補としての部分集合 $M_1, M_2, \dots, M_n$ に分割されているとする。同定のための作業空間を V とし、 $M_i (i = 1, \dots, n)$ に関する認証のための作業空間を W_i とする。集合 M に属する2つの元 p, q について、 V および W_i における距離をそれぞれ $d(p, q)$ および $d_i(p, q)$ と表す。

V において距離尺度の意味で部分集合 M_i を代表する1つの元(代表元)を m_i とする ($m_i \in M_i$)。また、同様に W_i における M_i の代表元を r_i とする (一般に $m_i \neq r_i$)。

さて、不明個体 x が与えられたとき、 x を同定するもっとも簡単な手順は、作業空間 V において x と n 個の代表元 $\{m_i\}$ との距離 $\{d(x, m_i)\}$ の中で最小値をしめす代表元 m_k をみつけ、 x が部分集合 M_k に属すると判定する方法である。この場合、どういう作業空間 V を設定するかによって同定の成否がきまるといっても過言ではない。問題の根源は、「個」の候補たる各部分集合がそれ自身作業空間の中で一定の広がり(分布)をしめす中で、ちがった部分集合をきちんと識別できる座標軸をうまく設定できるかどうかにある(図2参照)。1つの部分集合が代表元を中心として作業空間の中で、ど

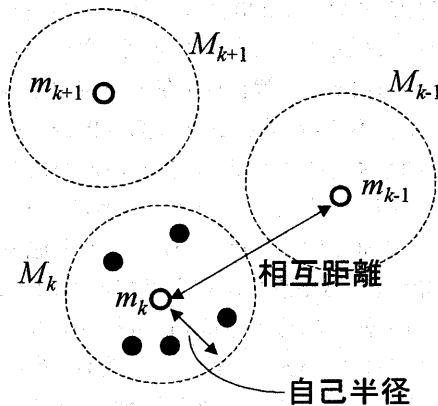


図2 作業空間

のぐらいの広がりをもつかを距離尺度ではかることができるとして、これを自己半径とよぶことにする。一方、ちがった部分集合の間の距離はできるかぎり大きいほうが望ましいわけであるが、これのめやすとして代表元の間の相互距離に注目する。ここで、「個」の同定にあたって、望ましい作業空間をどう設定すればよいかの目標をつぎのように要約する。

- ◆ 自己半径をできるかぎり小さく、相互距離をできるかぎり大きく、座標軸の数はできるかぎり少なくすること(効果的な座標軸の選択)。

つぎに、「個」の認証について考えてみる。認証は、たとえば $x \in M_k$ という仮説を検証する操作であって、この場合の作業空間 W_k の設定は、 M_k とその補集合 ($M - M_k$) を効果的に識別できる座標軸をどう設定できるかにかかっている。目標とするところは同定と同じであるが、2分類である点にちがいがあ

さて、上記の目標を達成する手順や具体的手法を考える。最初にやらなければならないことは座標軸の設定である。前述のように、「個」が属する集合の元たちを観察する視点としていくつかの属性を導入すれば、それぞれに対応する座標軸が生まれる。ここで、どういう属性に着目するかは、究極のところ、われわれの判断(直感+経験)に帰着する。いったん「個」を観測する属性を決めると、同定や認証の基礎になる情報がそれによって決まってしまう点に注意が必要である。つまり、「個」の根源たる固有情報をもれなく捕捉できるような属性たちをみつけられるかどうか重要である。情報が欠落していると、いく

ら後の操作を工夫してみても所詮満足な結果を得ることはできないからである。こうした点を考慮しつつ、望ましい座標軸の設定に到達する現実的な手法は概略つぎの2つに大別できるであろう。

直感的手法 属性の導入からはじまって作業空間の設定にいたる全過程をすべて試行錯誤的・直感的に行う手法。いうなれば名人芸的手法。

数理統計的手法 属性の導入は直感に依存するとしても、その後の過程はできる限り数理的・統計的な考え方を援用していく手法。この場合、属性の導入にあたっては、考えるかぎり多数の属性をひとまず導入しておいて、数理的・統計的方法によって各属性の重要度を評価し、重要度の高い属性を選択的に残していく手法など種々のアプローチが考えられる。



図3 例題集合

3. 実験的検討

3. 1 例題集合と属性

「個」の同定と認証を考えるために、ここで3人の浮世絵作家（喜多川歌麿・葛飾北斎・安藤広重）の美人画から切り出した18個の顔のコレクションを例題集合 M とする（図3参照）。なお、番号1～6は歌麿、7～12は北斎、13～18は広重である。ここで、「個」の同定と認証の例題をそれぞれつぎのように設定する。

【同定】与えられた顔の作者は、喜多川歌麿か、葛飾北斎か、あるいは安藤広重か。

【認証】与えられた顔の作者が北斎かどうか、それは本当か。

例題は、3人の浮世絵作家の顔（美人画）の描き方にはそれぞれ個性があり、顔をみれば作家が誰かをおよそ判定できそうな傾向があることに注目して設定したものである。顔の描き方という観点でみれば、輪郭の取り方や目鼻口などの部品の配置状況として作家の

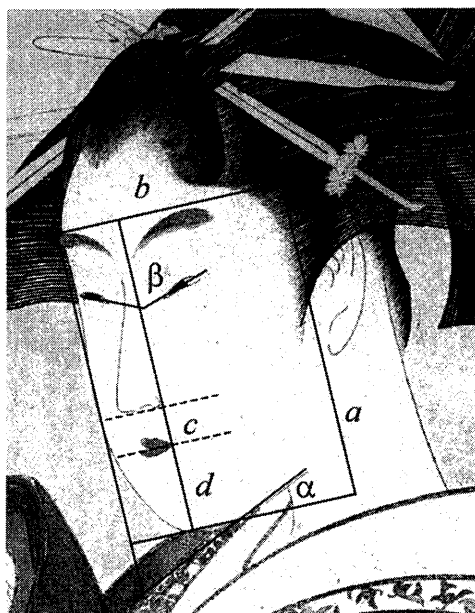
固有情報が捕捉できると考えられる。そこで、つぎのような各属性を導入する(図4参照)。

〈輪郭縦横比〉矩形型輪郭の縦横比 (z_1)

〈口唇点内分比〉口唇点が鼻-顎線分を分割する内分比 (z_2)

〈顎角〉顎角の正接 (z_3)

〈両眼交叉角〉両眼の交叉角の正接 (z_4)



$$z_1 = b/a \quad z_2 = c/(c+d)$$

$$z_3 = \tan \alpha \quad z_4 = \tan (\beta - 90^\circ)$$

図4 4つの属性

表1 計測値一覧

顔番号	Z1	Z2	Z3	Z4
1	0.73	0.33	0.32	1.00
2	0.70	0.32	0.26	0.80
3	0.69	0.31	0.21	1.08
4	0.72	0.36	0.29	0.96
5	0.74	0.30	0.33	0.86
6	0.73	0.33	0.43	1.08
7	0.66	0.43	0.52	1.29
8	0.64	0.44	0.65	1.27
9	0.67	0.42	0.61	1.14
10	0.67	0.67	0.77	1.07
11	0.68	0.68	0.53	1.00
12	0.66	0.66	0.63	0.96
13	0.62	0.31	0.32	0.26
14	0.62	0.28	0.47	0.39
15	0.63	0.34	0.53	0.35
16	0.61	0.32	0.30	0.45
17	0.63	0.30	0.43	0.47
18	0.62	0.30	0.22	0.57
平均値	0.67	0.33	0.43	0.83

例題集合に属する個々の顔(番号1~18)について、上記の4つの属性を計測して得られた計測値を一覧表(表1)に示している。

3. 2 判別分析による同定

判別分析では、与えられた集合をもっとも効果的に所期の部分集合へと分割できる座標軸系の設定を追求する。ここでは、例題集合を歌麿グループ、北斎グループ、広重グループの3つの部分集合へと効果的に分類できる新しい座標軸(属性値の1次結合)をみつけることが問題になる。今回の実験的分析にあつては、簡便に既製統計ソフト(STATISTICA)を利用することとした。この結果、次式で示される2つの新しい座標軸 Q_1 および Q_2 が例題集合の分類に有効であることが判明した。

$$Q_1 = -41.21z_1^* - 8.06z_2^* + 0.34z_3^* - 8.86z_4^* \quad (1)$$

$$Q_2 = 60.75z_1^* - 8.88z_2^* - 6.78z_3^* - 3.11z_4^* \quad (2)$$

ただし、 $z_1^* \sim z_4^*$ は、 $z_1 \sim z_4$ からそれぞれの平均値を差し引いた値である（表1参照）。

さて、2つの新しい座標軸によって2次元の作業空間がつけられることになるが、この空間における例題集合の分布状況を図5に示す。3つのグループが鮮明に分離されていることがわかる。与えられた「個」 $x \in M$ の同定にあたっては、まず x の属性を計測し、計測値から式（1）および（2）にしたがって Q_1 および Q_2 を算出する。この後、作業空間において x がどのグループの位置を占めるかを判定すればよい。

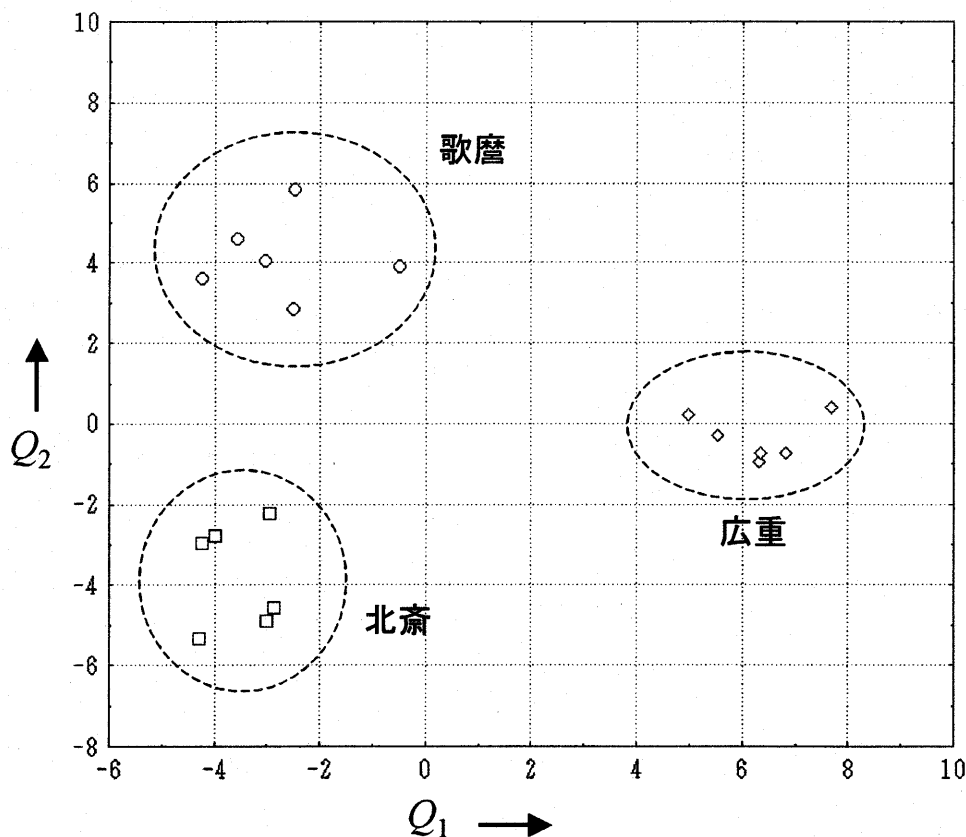


図5 作業空間における例題集合の分布

3. 3 判別分析による認証

ここでの認証は、前述のように北斎画の認証である。例題集合において北斎の部分集合（7～12）を「個」とし、歌麿と広重（1～6および13～18）を他者とした2分類を考える。ここでも、判別分析を行う。この結果として、次式で示される1つの新しい座標軸によって効果的に2分類できることが判明した（表2参照）。

$$R = 36.05z_1^* - 11.95z_2^* - 6.04z_3^* - 7.05z_4^* \quad (3)$$

ただし、 $z_1^* \sim z_4^*$ は、式 (1) および (2) の場合と同様である。

ここで、与えられた「個」 $x \in M$ が北斎であるという仮説について、認証のための判定基準を具体的に示すとすれば、たとえば、表 2 を参考にしてつぎのように書くことができるであろう（厳密には統計的検定が必要になる）。

- ◆ $R \leq -3$ ならば、仮説は真である
- ◆ $R > -3$ ならば、仮説は偽である

4. むすび

本稿では、「個」の同定と認証について分類の視点からの一考察を述べた。とくに、同定と認証という機能を情報处理的に解釈した場合、それぞれの作業空間が異なってしまうべきであるという点を強調した。実験的検討には、同定と認証についてともに 3 人の作家による浮世絵の顔の例題集合を用いたが、認証の実験的検討としてはむしろ別の集合を用いた方が適切であり、实际的であろうかと思われる。つまり、認証が必要な具体的な場面として、たとえば真贋判定を考えることができるが、この場合の例題集合は北斎の真作と偽作の集合でなければならない。実際にこうしたデータが入手できれば、実験的検討を行ってみたいと思う。

いずれにせよ、「個」の同定と認証においてもっとも根源的な要件は、集合の元たちを観察する視点としていかに効果的な属性をみつけられるかにかかっている。これは、まさにわれわれの直感力に依存するところが大きい作業であって、直面する問題ごとに個別に対応する必要がある。これについての一般的な方法論はみいだし得ないと考えられる。このほか、紙面の都合で割愛した検討項目も多い。稿を改めて詳述したいと考えている。

【参考文献】

- [1] 柳井・高根、『多変量解析法』，朝倉書店，1977.
- [2] I.T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, 1986.
- [3] 小沢，分類における重複性の表現手法—重複クラスタリング—，情報処理学会『人文科学とコンピュータ』研究会，34-1，1-8，2002.
- [4] 檜崎，『大英博物館所蔵 浮世絵名作展図録』，朝日新聞社，1985.
- [5] 狩野，『広重』，平凡社，1984.
- [6] 永田，『北斎』，平凡社，1984.

表2 算定値

顔番号	R
1	2.53
2	3.34
3	1.43
4	2.27
5	4.18
6	1.30
7	-4.44
8	-5.92
9	-3.45
10	-6.91
11	-4.72
12	-5.53
13	4.02
14	2.56
15	2.12
16	2.32
17	2.35
18	2.56