

図 1 ヴァーサ号博物館の展示(撮影:佐々/蘭/

ユネスコの水中文化、 一保護条 为でた、「水中遺跡の管理方法を検討する際、遺跡の現状保存を第一の 選択 人、する」ことに主眼が当てられている^[9]. 水中遺跡は、放置しておくと気化することも考えられ、遺跡を守るために時には遺跡を完全に埋めたす手にも取る。一見すると、「水中遺跡は発掘しない」がスタンダードであるとこと思えるが、効率よく遺跡を調査し、その成果を広く一般へ公開することが求められている。

2. デジタル技術の革新がもたらすもの

効率の良い遺跡調査を行うためには、デジタル技術の応用は重要な鍵となる. 表1に水中遺跡の調査と活用の主な問題点、それらを解決するために必要なデジタル技術をまとめた. 以下、近年のデジタル技術の発展がもたらしている調査手法の変化とその例を示す.

水中遺跡特有の問題点	デジタル技術の応用	
遺跡の位置の特定が困難	GPS による位置特定 GIS によるデータ管理と公開	
遺跡の悉皆調査が困難	水中探査機器の効率向上(計算処理速度) 水中探査機器の小型化など マルチビーム・ソーナーなど3次元測量が可能	
水中での測量 (時間的制約など)	デジタル 3 次元測量 (潜水時間の短縮) 高感度・高画質, 情報共有の効率向上	
保存処理が困難 (主に金属製品)	金属製品の CT スキャン 保存処理の確認	
遺跡へのアクセスができない (活用が困難)	3D プリンターなどによる、オラー 夏元 VR や AR	

表 1 水中遺跡の調査と活用の主な問題点とデジタル技術の応用

2.1 悉皆調查: GPS/GIS

遺跡の調査において、その位置だ特定上プことは基本的な作業である。海の上における GPS の利便性については明、かであるが、位置情報を管理するデータベース(GIS)の存在な上きい。 ジンマークでは、石器時代の水没遺跡だけで 7,000 件を超っており、他国ごも同様に数千から数万の水中遺跡が知られている(図 2)。 の 、 に 版大な量のデータは、GIS による管理が望ましい i 。 20($^{\circ}$ 年から Eに 諸国を中心とした MACHU(Managing Cultural Heritage Underwater) Project では、文献史料に見られる沈没の情報、漁師などによる引き掲が近初の報告、周知の遺跡の位置、遺跡の時代や特徴など様々な情報を一括で管理する GIS データベースを作成し、水中遺跡の管理の方法を提示した $^{[6]}$.

2.2 探查技術

水中遺跡を発見するには、水中を見る方法(水中探査)が必要となる(図3)、水中探査に利用する機器には、その原理も音波を用いるものなど様々ある(表2)、海底の環境、遺跡の特徴、調査目的や予算規模によってどの探査機器が最も有用であるか(例えば、金属を使用していない古代の木造船を発見するには金属探知機は使えないなど)を判断し、また、データを分析する「目」も

3. Twitter のログデータからわかる地域の諸文化

Twitter のログデータは、Twitter 社が提供する **API** (application programming interface)を通して誰でも取得できる。Twitter のログデータには、様々な文化情報が含まれており、方言や食文化に関する分析資料としての利用可能性が期待されている^[3]。例えば、日本国内の位置情報が付与された Twitter のログデータのうち、「うどん」を含むものは主に西日本で卓越し、特に香川県や岡山県、愛媛県、福岡県で多くなっている(図 1). Twitter のログデータには、投稿時間の情報も含まれており、東京都、大阪府では、昼食と夕食の時間帯に「うどん」を含むログデータが多い、方、香川県では朝 7時台をはじめとして、朝食と昼食の時間帯に「うどん」を含いログデータが多い傾向にあり、朝食にうどんを食べる香川県の文化、物特徴が現。、いる^[3].

また、Kirimura [4] は、Twitter のログ ータ こ名 たれ 日常会話に注目し、「(笑)」や笑顔の顔文字を含む投稿 抽品 し 地域別に集計した。西日本では「(笑)」が多く、東京周辺では「(笑)」、大笑 「の顔文字もそれほど多くはない傾向が確認された(図 2)。 現 大日本の文 この発信地である東京周辺では、より多くのパターンの顔、「や感情を現。生み出されているためと考えられる。

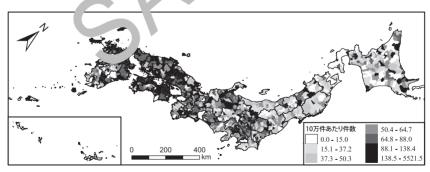


図 1 「うどん」を含む Twitter ログデータの地理的分布(桐村^[3]より引用)

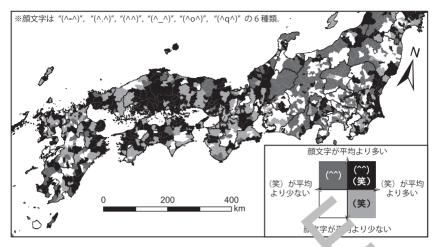


図 2 笑顔の顔文字・「(笑)」を含む Twitter ログデータの件。ケに封っく地域分類 (Kirimura^[4]より引用・編集)

4. 名字データからわかる地域(1) 歴 つ的結び付き

日本においても、中谷らが「日本人の名字マップ」をウェブ公開している^[1]. また、桐村^[5]は、1,600 万件の電話帳データを利用して全国の名字がどのような地域で多いかを集計して名字を分類し、地域ごとに特有の名字のグループを求めた。この名字のグループ別に名字の件数を集計して、その構成比に基づいてさらに地域を分類した結果が図3である。この図からは、北海道の南部と東北地方の北部(R04)、北海道の大部分と北陸地方(R05)など、明治以降の開拓に伴う地域間の歴史的な結び付きを読み取ることができる。



図 1 国会図書館デジタルコレクション・奈良絵本 「ぇ くさ太´´」」 IIIF によるページ問覧.

位でのアノテーションや画像の比性なとも「能となってきた、今後は、IIIFの環境を使った閲覧や情報学的研究」、流といるであろう(図1).

4. 研究の可能性

絵画資料を「象とした場合、対象によっては、その大きな面積の画像を、限られた大きさのニーター上にどのように再現するかという課題がある。基本的に、データが送出されてくるのは、インターネット回線であり、高速にストレスなく、拡大縮小、全体表示・部分表示が可能となり、原物を目の前にしている以上に精細に観察できるような仕組みが必要である。地図のデジタル閲覧などと同様、ピラミッド画像を使って、絵巻全体をデジタルファイルにして配信する方法をとることが多いが、この方法は、一方で全体画像をダウンロードできなくするという副次的な機能を獲得することができる。そのため、コンテンツの悪用・流用などの目的でのダウンロードを避けたい所蔵者が情報配信するための方法として使われる。オープンデータ化の流れが主流となる中で、デジタルファクシミリでのWeb上での一般公開は、人文学的な研究には一歩前進にみえるが、こうした制限付きの公開方法では、デジタル環境下で加速される

文化情報学的な研究方法を遮断してしまう. オープンデータ化することで得られる, 思ってもみなかった活用方法へとつなげられる機会を失っていることも, 視野に入れておく必要があろう.

このような単純なデジタル閲覧は、結局、原物に直接アクセスしての熟覧を 越えるものではない、一方で、文化情報学型研究は、デジタル化されたがため に可能となる研究に本領がある。現在、次のような研究が目立つものである。

① **画像マッチング**:描かれている画像の色や形態を解析して類似画像や同一画像を抽出.

とくに版画の分野において同一画像を瞬時に収集することで、検索の効果を 各段に上げることに成功している(図 $\mathbf{2}$) $^{[2]}$. また、浮世、の美人を使った顔 の表現を計量的に分析し、絵師の特徴を抽出し、絵師、特定に利用する研究も 行われている $^{[4]}$.

② メタデータの活用:異言語によって記述 これた複数の データベースへの 並列アクセスによる作品の同定・検索や 検索 青果 推 ボーステム [3].

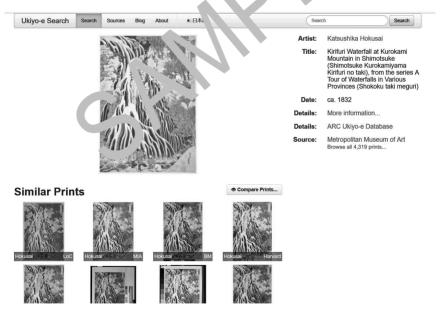


図2 Japanese Woodblock Print Search (Ukiyo-e.org) 画像マッチングによる検索,画像比較が可能.

A5 文化・芸術



図 1 ARC 古典籍ホータル ータへ 人

コンテンツを保有する機関とたい。これ、のメタデータを集約するポータルサイトであり、当然そいこには書かに、する情報が大量に集積され、諸機関を横断した検索や閲覧が可能によった述の IIIF にも対応しているほか、大量のコンテンツ いら得られた知識を再構成し、発信できるギャラリー機能も備わっており、連接の異の場大が期待される。

6. 古典籍デジタルアーカイブの展望と課題

ARC のポータルデータベースにせよ、IIIF にせよ、所蔵機関の枠を超えて、書物や各分野の関連資料を有機的に結び付けていこうとする方向性は一致している。現在は、古典籍デジタルアーカイブは、古典籍を Web 上で閲覧する目的だけでなく、その上で書物学や出版研究を行うツールとして姿を現しつつあるといえよう。その間口を広げるためには、画像のみならず、翻刻テキストの整備を徐々に進めることが必要と考えられる。翻刻とその校正には多大な労力を要するが、近年、京都大学古地震研究会の共同翻刻アプリケーション

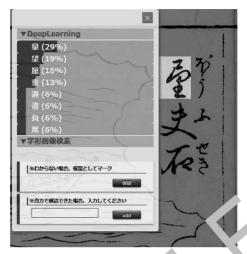


図2 翻刻支援システム(ARC 古典籍ホータルデータ * 人)

古典籍デジタルアーカイブは、今後も量の充実が図られ、書物学や出版研究 に資するところがさらに大きくなるだろう。また、ARCのポータルデータベ ースや IIIF を含め、新規技術開発というよりはむしろ運用面において、今後

$$F = \frac{\frac{\chi_x^2}{n_x - 1}}{\frac{\chi_Y^2}{n_Y - 1}}$$

が、自由度 $n_X - 1$ 、 $n_Y - 1$ の F 分布(F distribution、付表 4)と呼ばれる分布に従うことを利用する.

カイ 2 乗分布, F 分布は左右対称の分布ではなく,右にスソを引く分布になる。

[村上征勝]

【参考文献(さらに学びたい人のために)】

- [1] ホーエル、P., G. (村上正康、浅井晃訳) (1981)、『初等統計与 ' ヴュ館、
- [2] ジョンソン, M., K., リーバート, R., M. (村上征牒, 西平重喜, ¹) (19²). 『統計の基礎』 サイエンス社.
- [3] 松原望(2007). 『入門統計解析』東京図書.
- [4] 鄭躍軍, 金明哲, 村上征勝 (2007). 『デーマサイエンス入 『 心誠出版.

B1-5

推定と検定

statistical estir and tes

1. 点推定,区型地足

母集団の平均値、分散、比率などのパラメーターを、n 個の標本から得られた 1 つの値で推定する方法を**点推定**(point estimation)という。母平均を推定するには標本平均値 \overline{X} が、母分散 σ^2 を推定するには標本の分散 s^2 が、母比率 p を推定するには標本比率 \hat{p} が用いられる。

一般に母集団のパラメーターを推定する統計量には、不偏性、一致性、充足性、有効性という 4 つの性質を有することが望まれるが、 $\mathbf{B1-3}$ で定義した標本分散 s^2 は**不偏性**を持たないので、その代わりに不偏性を持つ**不偏分散**と呼ばれる次の統計量

$$s^{2} = \frac{1}{n-1} \sum_{i=1}^{n} (x_{i} - \overline{X})^{2}$$

が用いられることが多い.

このように、分散の定義は2種類あるが、分析にどちらを用いても分析結果は同じになる。ただ分析途中の数式は多少異なるので、統計の本を使うときは、その本の著者がどちらを用いているか事前に注意しておく必要がある。

さて、母集団の平均値、分散、比率などのパラメーターが、ある確率 $100 \cdot (1-\alpha)$ % で入る区間を推定するのが**区間推定** (interval estimation) である.

母平均 μ の $100 \cdot (1-\alpha)$ % 信頼区間は、分散 σ^2 がわかっている場合は、標準正規分布の右すその面積が $\alpha/2$ % となる Z の値を $z_{\alpha/2}$ とすると

$$\overline{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

となる.

分散 σ^2 が不明で不偏分散を用い、場上'は 自由度 n-1 の t 分布の右すその面積が $\alpha/2$ % となる t の値を $t_{\alpha/2}$ と たる

$$\overline{X} - t_{\alpha/2} \sqrt{\frac{s}{2}} < \epsilon < \overline{X} + t_{\alpha/2} \frac{s}{\sqrt{n}}$$

となる。

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

母比率 p の $100 \cdot (1-\alpha)$ % 信頼区間は、標本比率を \hat{p} とすると

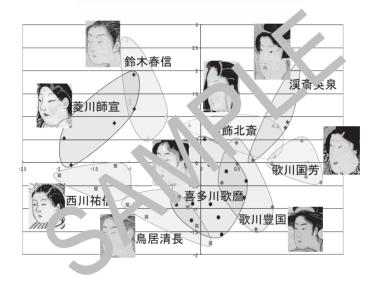
$$\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{m}}$$

となる.

2. 仮説検定

母集団のパラメーターに関し、データを得る前に仮説が設定できる場合

量的分析法



浮世絵美人画の主成分分析

9人の浮世絵師の描いた女性の顔を主成分分析し、形の似た顔が近くになるように配置したグラフ. このグラフでは、活躍期の遅い絵師ほど右側に位置する結果となったが、グラフの横軸は顔の長さに関係した変数であり、右へいくほど顔は面長になることから、江戸時代の庶民の好む女性の顔は、丸顔から面長な顔に変わっていったことがわかる.

B2 量的分析法

例えば作品の作者名や性別というような個体を区別するためにつけられたラベルのことである.

そのような名義尺度を利用した主成分分析ではどのような知見が得られる可能性があるかを,**浮世絵美人画**の分析例で示す.

2. 浮世絵美人画の分析

分析対象の浮世絵は、江戸時代の代表的な9人の浮世絵師、菱川師宣、西川祐信、鈴木春信、鳥居清長、喜多川歌麿、葛飾北斎、歌川豊国、渓斎英泉、歌川国芳の描いた女性の顔53点である。表1は、9人の絵作の活躍期と分析に用いた顔の数である。なお9人の絵師の活躍期を便宜した前期・中期・後期の3つに区分しておく。

分析に用いた顔の変数は、目、鼻、口、耳なこの顔の部位 、箇所の位置座標から求めた、図 1 に示した 12 種類の 1 度で 5 る 2 0 12 種の角度情報に対して相関行列を用いた主成分分析を行った。 電 2 である。図 2 の横軸は第 1 主成分、縦軸は第 2 主成分で、 2 1 成分までの累積寄与率は 0.56 であるので、わずか 2 個の合成 3 つ中に 5 。個の顔に関する 12 種類の変数の情報の約 56% がこの図に、まれていることになる。図中の 53 個の丸印はそれぞれ分析に用いた顔を表 、 1 1 音類の角度が類似した顔は近くに位置している。

この図の曲 *で囲まれた範囲は、描いた絵師の名前のラベル(名義尺度)を

活躍期の区分	絵師	活躍期	分析した顔の数
前期	菱川師宣	1670~1694	5
	西川祐信	1692~1750	6
	鈴木春信	1760~1770	5
中期	鳥居清長	1767~1815	6
	喜多川歌麿	1775~1806	11
	葛飾北斎	1779~1849	4
	歌川豊国	1788~1825	5
後期	渓斎英泉	1810~1848	6
	歌川国芳	1812~1861	5

ま 9人の絵師の活躍期と分析した顔の数





変数 8: ∠33, 1, 29 変数 10: ∠6, 0, 2 $12: \angle 1, 10, 12$ $13: \angle 10, 0, 12$

 $11: \angle 7, 0, 3$ $17: \angle 1, 17, 0$ 21: $\angle 1$, 26, 21

変数 2: ∠32, 0, 28 $20: \angle 1, 16, 0$ $27: \angle 1, 10, 11$

 $16: \angle 1, 17, 21$

 $23 : \angle 1, 25, 21$

図 1 分析に用いた 12 種類の角度

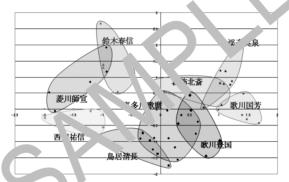


図2 美/ 国の主成分分析(相関行列)(B2 扉図参照)

用い、同一の絵師の描いた顔の範囲である.

この図2から、少なくとも次の3つのことを知ることができる.

まず、この図でそれぞれの絵師の描いた顔の範囲を曲線で囲ってみると、9 人の絵師の描いた顔の範囲は比較的狭く、かつ、他の絵師と重複する部分が少 ない. したがって、わずか 12 種類の角度ではあるが、9 人の絵師の描く顔の 特徴がある程度把握できていることがわかる.

また9人の絵師の描いたすべての顔の類似関係が俯瞰できるため、9人の中 で誰と誰の描き方が類似しているのかを、全体を俯瞰し視覚的に把握すること