

# 花押を対象としたデータ駆動型歴史情報学研究の実践

中村 覚・劉 冠偉・宮崎 肇・井上 聡 (東京大学 史料編纂所)

大山 航 (東京電機大学 システムデザイン工学部)

山田 太造 (東京大学 史料編纂所)

**概要**：東京大学史料編纂所では、「データ駆動型歴史情報研究基盤の構築」プロジェクトを実施している。本研究では当該プロジェクトにおける1つの適用事例として、花押を対象とした歴史情報学研究の実践例について述べる。具体的には、花押の検出や分類など、花押に対する定量的な分析を試みることで、データ駆動型の歴史情報研究のモデル検証を行う。

**キーワード**：古文書学，花押，歴史情報学，機械学習，可視化

## Implementation of data-driven historical informatics research on Kao Signature

Satoru Nakamura / Guanwei Liu / Hajime Miyazaki / Satoshi Inoue (Historiographical Institute, The University of Tokyo)

Wataru Ohyama (Department of Information System Engineering, Tokyo Denki University)

Taizo Yamada (Historiographical Institute, The University of Tokyo)

**Abstract**: Historiographical Institute, The University of Tokyo has been implementing the "Construction of Data-driven Historical Information Research Infrastructure" project. This study describes a practical example of historical informatics research on Kao (signature) as one of the applications of this project. Specifically, we attempt quantitative analysis of Kao, such as detection and classification of them, to verify a model for data-driven historical information research.

**Keywords**: Paleography, Signature, Digital History, Machine Learning, Visualization

### 1. はじめに

東京大学史料編纂所は明治以来 150 年に渡り蓄積してきた目録、画像、本文、文字など大量かつ多様なデータを有している。具体的には、1,200冊近い基幹史料集の編纂・刊行実績、40種560万件のデータベースと史料画像2,000万件のデジタルアーカイブを持つ。これらの豊富なデータと、50名以上の所内研究者をはじめとする専門家が持つ知識を有機的に繋げることにより、歴史情報の発信を支える新たな基盤の構築を目指している。この一環として、東京大学史料編纂所では、2021年度より、「データ駆動型歴史情報研究基盤の構築」プロジェクトを実施している[1]。

本研究では当該プロジェクトにおける1つの適用事例として、花押を対象とした歴史情報学研究の実践例について述べる。具体的には、花押の検出や分類など、花押に対する定量的な分析を試みることで、データ駆動型の歴史情報研究のモデル検証を行う。

関連する研究として、北本[2]らによる「江戸マップ」や「顔貌コレクション」などが挙げられる。本研究では、これらのデータとの互換性を保持しつつ、東京大学史料編纂所が有する多種多様なデータを取り扱う点に違いがある。また、中村らは図1に示す「データ駆動型歴史情報研究基盤の構築手法」を提案し、絵図史料への適用を行って

いる[3]。本研究では、同手法をベースとしつつ、適用対象を花押データとして、各種可視化や分析を試みる。

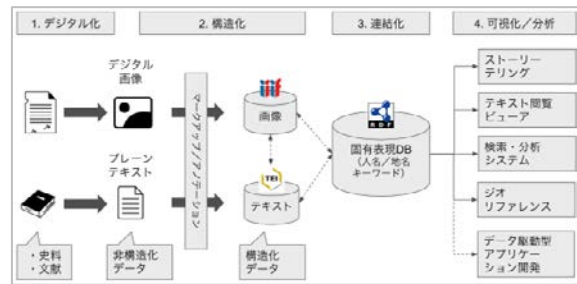


図1 データ駆動型歴史情報研究基盤の構築にむけた提案手法

Figure 1 Proposed methods for constructing a data-driven historical informatics research infrastructure.

### 2. 花押の研究例

花押を対象にした歴史学・古文書学に関する研究は、これまで数多くの研究成果を生み出している。例えば、佐藤進一による『[新版]古文書学入門』[4]や『増補 花押を読む』[5]、上島有による『中世花押の謎を解く一足利将軍家とその花押一』[6]などが挙げられる。これらの中で、花押の

位置の変化は署判者の政治的位置の高下や受給者との力関係を直接的に反映するものである、と述べられている。また、花押の大きさについても、足利尊氏・直義花押の大きさは署判位置と連動するかたちで地位が上がるにつれ巨大化する傾向にある、と述べている。このように、花押の位置と大きさが文書の書札礼の重要な要素であり、発給者の地位の高下や受給者との関係性をあらわす重要な情報であることが、古文書学・日本史学の共通認識となっている。ただし、このような共通理解に対するデータが十分に示されていない点が課題の一つとして挙げられる。

これらの既存研究に対する本研究の差異は、画像処理技術等を用いた定量的な分析・データ駆動型の研究手法を適用する点である。これにより、既存の研究成果に対する定量的なエビデンスのサポートや、人手を中心とした方法では発見が難しかった新たな研究課題を創出することを目的とする。画像処理技術等を用いた花押に対する既存の研究としては、鬼塚らによる研究[7]や岡野らによる研究[8]がある。鬼塚らは画像をクエリとしたデータベース検索を実現するために、機械学習手法のひとつである畳み込みオートエンコーダを用いて花押の字形特徴を抽出し、字形特徴の相違度を算出する手法を提案している。また岡野らは IIF を活用し、複数の花押画像を対照表示できる機能を実装している。

本研究では、画像処理技術等を用いた定量的な分析にあたり、古文書学の手法をベースとする。古文書学では、例えば文書に使用される紙を「料紙」と呼ぶ。また書状の形式に関する料紙の使用方法は、堅紙（たてがみ）やその略式の書状である折紙（おりがみ）などに分類される。図 2 に折紙の例を示す。

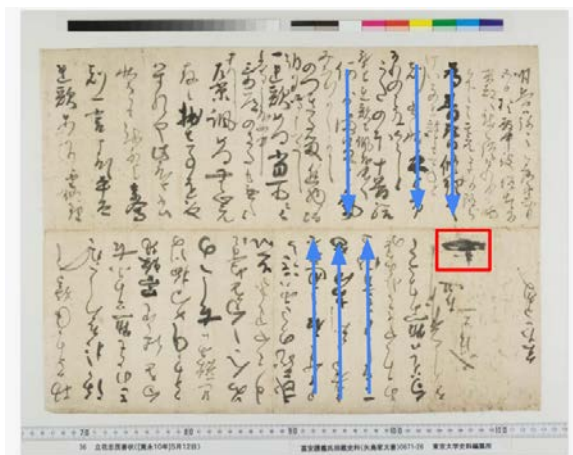


図 2 「折紙」の例

Figure 2 Example of Origami (Folding paper).

さらに花押をはじめとするサインの位置についても、図 3 に示すような分類がなされている。

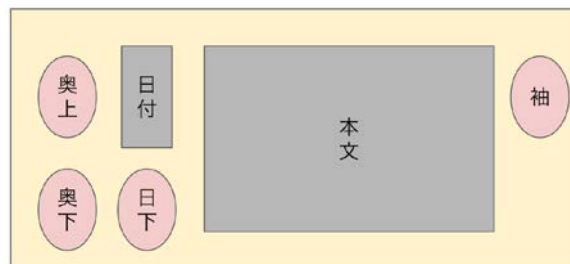


図 3 位置による花押の分類

Figure 3 Classification of signature by position.

具体的には、文書の右端の余白部分である「袖」、日付の次行上部である「奥上」、日付の次行下部である「奥下」、および日付の下部である「日下」などに分類される。袖判は宛所に対してもっとも尊大な書札礼となる。一方、奥上署判の文書は、日下・袖判の文書に比べて、相手に対して丁寧な様式となる。

本研究では、このような古文書学の手法に基づき、東京大学史料編纂所の花押データベースの史料画像およびメタデータを対象として、画像処理技術等を用いた定量的な分析を試みる。花押データベースには、史料の作成年月日、史料画像中の花押の位置、およびその人物に関する情報等が蓄積されている。これらのデータに対する探索的データ分析 (Explanatory Data Analysis: EDA) を試みることで、先述した通り、既存の研究成果に対する定量的なエビデンスのサポートや、人手を中心とした方法では発見が難しかった新たな研究課題を創出することを目指す。

### 3. データ作成

花押に対する探索的データ分析に向けて、既存のデータの整形、および分析に必要な追加データの作成を行う。なお、後述するデータ整理のプロセスについて、画像に対するアノテーションとして行うことにより、本プロセスの結果を機械学習のためのデータセットとしても利用可能とする。

#### 3.1. 花押データベースからのデータ取得

まず、東京大学史料編纂所の花押データベースから、史料画像と、その画像中の花押の位置などに関するデータ（以下、アイテムと呼ぶ）587 件を取得した。アイテムのメタデータ項目と値の例を表 1 に示す。

また、これらのアイテムが格納された史料群の上位 5 件の名称とデータ数の内訳を表 2 に示す。小笠原文書や島津家文書 (歴代亀鑑・宝鑑) など、代表的な武家文書から抽出した花押のデータを対象とする。データ数は限られるが、これらのデータから得られる傾向は、当時の典型的な例を示すと考えられる。

表 1 アイテムのメタデータ例  
Table 1 Example of item metadata.

項目	値の例 1	値の例 2
人名	六郎別当神主	有千代大夫
人名注記		
備考		1-18-21
史料群名	菅浦文書	賀茂別雷神社文書(II-B-1)
和暦年月日	年未詳	明応4年11月20日
文書名	菅浦惣庄乙名置文	氏人中起請文
位署書	六郎別当神主 K	有千代大夫

表 2 分析対象とする史料群の例  
Table 2 Example of target historical documents.

文書名	史料群名	データ数
小笠原文書	小笠原文書	258
島津家文書	歴代亀鑑(五三通)	66
島津家文書	宝鑑其一(五一通)	61
島津家文書	宝鑑其二(五三通)	59
島津家文書	歴代亀鑑(五四通)	53

### 3.2. 料紙範囲の検出

3.1 で取得したデータについて、「料紙」領域の抽出を行った。図 4 の青枠で示すように、一部の画像では複数の料紙が含まれる場合がある。画像を単純に左右に 2 分割した場合、画像中の花押の相対位置と、料紙中の花押の相対位置に違いが生じるため、料紙領域の抽出が必要となる。

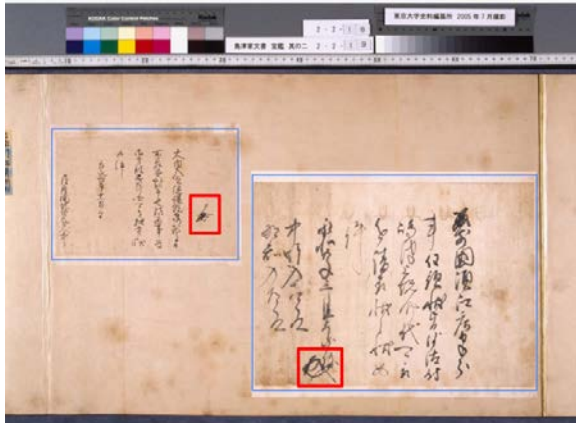


図 4 画像に複数の料紙が含まれる例  
Figure 4 Example of an image that includes more than one paper.

### 3.3. 文書形式の分類

図 2 に示した「折紙」などの場合には、料紙内の花押の位置を単純には算出できない。そのため、画像中の料紙を「堅紙」「折紙」に分類するメタ

データを人手で付与した。この結果、表 3 に示す分類結果となった。

表 3 「堅紙」と「折紙」の分類結果  
Table 3 Classification Results of “Tategami” and “Origami”.

分類	画像数	花押(アイテム)数
堅紙	376 件	547 件
折紙	36 件	40 件

折紙については、図 2 に示したように、花押の位置に関する分析が、堅紙に比べて困難である。そのため、本研究では後述する過程において、折紙を分析対象から除外している。また本研究では切紙(きりがみ)も「堅紙」に分類している点に注意が必要である。後述する花押の位置や料紙に対する比率の計算に与える影響は小さいが、堅紙・折紙・切紙といった文書形式に基づく分析を行う場合には、堅紙と切紙の分類が求められる。この点は今後の課題とする。

### 3.4. 日付の位置の抽出

これまでのプロセスにより、料紙における花押の左右に関する相対位置を取得できる。次に、図 3 で示した花押の位置を「日下」「奥上」「奥下」に分類するため、日付文字列の位置の抽出を行う。その結果、412 件の文書画像から 421 件の日付の位置を抽出した。なお、3.2 で述べた通り、1 画像に複数の料紙が含まれる場合があるため、日付の個数が画像数よりも大きくなっている。一方、14 件の文書画像からは、日付の位置を特定することができなかった。

### 3.5. データのまとめと整形

これまでのプロセスで作成したアノテーションデータの例を図 5 に示す。

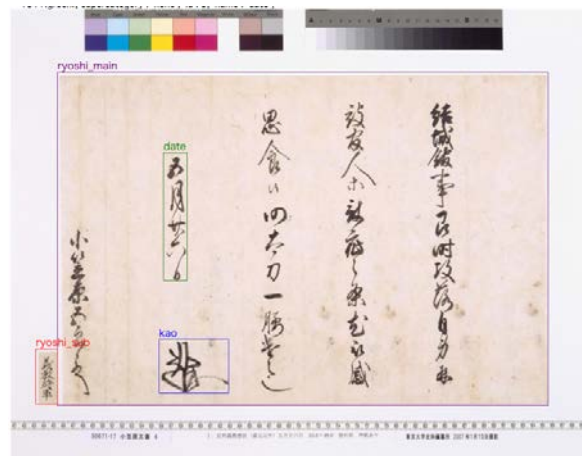


図 5 アノテーションデータの例  
Figure 5 Example of annotation data.

花押, 料紙 (のメインとサブ領域) および日付の位置をデータ化した. このデータを CSV などの形式にまとめ, 後述する探索的データ分析に利用可能な形に整形した. 特に, 史料画像のサイズ, および料紙, 花押, 日付の位置情報 (xywh 座標) を整理することで, 花押の位置情報や大きさに基づく分析を可能とする.

#### 4. ケーススタディ

本章では, 3 で作成したデータに対して探索的データ分析を行い, 花押に対する定量的な分析を試みる. 探索的データ分析とは, 数値の要約と視覚化を使用してデータを探索し, 変数間の潜在的な関係を特定するプロセスである. 本手法を用いることで, 古文書学・史料学における暗黙知の形式知化などを試みる.

探索的データ分析を行うツールとして, nbdev[9]を用いた. nbdev はソフトウェア パッケージや技術記事の作成, テスト, 文書化, 配布をすべて 1 か所の Jupyter Notebook で行う. 図 6 に nbdev によって GitHub Pages にデプロイされる文書の例を示す. 以下, 分析結果の詳細について述べる.



図 6 nbdev を用いた分析結果の文書の例  
Figure 6 Example of documentation of analysis results created using nbdev.

##### 4.1. 文書形式に関する分析

文書形式 (縦紙と折紙) の数量に関する時系列変化を可視化した結果を図 7 に示す. 50 年毎の花押数の数量を示し, そのうちの文書形式を色分けして示す.

図 7 から, 鎌倉時代 (1185-1333 年) には縦紙が使用されることが多いのに対し, 室町時代 (1336-1573 年) 以降に折紙が使用されていることがわかる. これは室町時代以降, 発給する文書

数の増加に伴い, 折紙 (や切紙) が使用され始める, という傾向を示すものと考えられる.

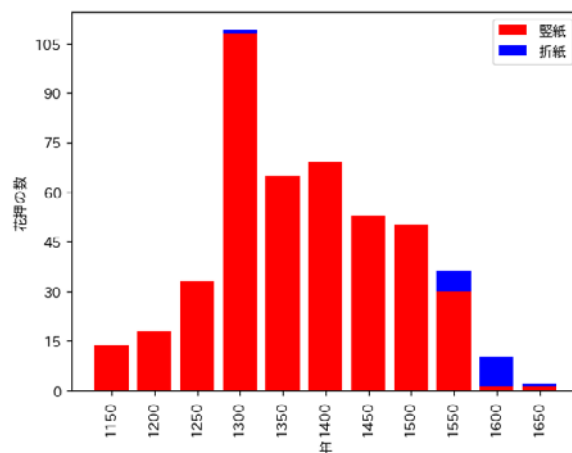


図 7 文書形式毎の花押数の時系列変化  
Figure 7 Time-series changes in the number of signatures by document format.

##### 4.2. 花押の位置に関する分析

3 章で作成した花押, 料紙および日付の位置関係に基づき, それぞれの花押を「袖」「日下」「奥上」「奥下」に分類した. 「袖」は料紙の右側 1/3 の領域に出現するものとした. 左上を起点とする xy 座標において, 「日下」は花押と日付の x 座標に重なりがあるものを分類した. x 座標が日付と重ならず, 日付より左側にある花押については, その y 座標から「奥上」および「奥下」に分類した. なお, 日付の位置が特定できなかったものなど, 一部の花押については「判別不能」と分類している.

図 8 に位置 (袖, 日下, 奥上, 奥下) 別の花押数の時系列変化を示す. また, 花押の数が多い人物の上位 20 人を抽出し, 人物および位置別の花押数を可視化した結果を図 9 に示す. また図 10 は同図を人物の生年で並び替えた結果を示す.

これらの結果から得られる特徴として, 「袖」に分類される花押が 1150-1200 年, および 1300-1450 年頃に集中していることがわかる. そして, 前者は初代鎌倉幕府の将軍である源頼朝, 後者は室町幕府の将軍である初代・足利尊氏, 2代・義詮, 3代・義満, 4代・義持による花押である. これは初期の室町幕府は袖判を多用するが, それ以降は使用しなくなる, という傾向を示すものと考えられる.

また, 室町幕府の 2 代将軍である足利義詮の文書の発給数が少ない理由は, 享年 38 と短命なことにとともに, 将軍在職期間も短いことが要因の 1 つとして考えられる. 具体的には, 初代尊氏の在職期間は 1338 年 8 月から 1358 年 4 月までの約 20 年間, 2 代義詮は 1358 年 12 月から 1367 年 12

月までの9年間、3代義満は1368年12月から1394年12月までの26年間となっている。将軍としての義詮の活動期は、尊氏の半分以下、義満の3分の1以下であり、文書発給数が少ない点も、これによるものと考えられる。

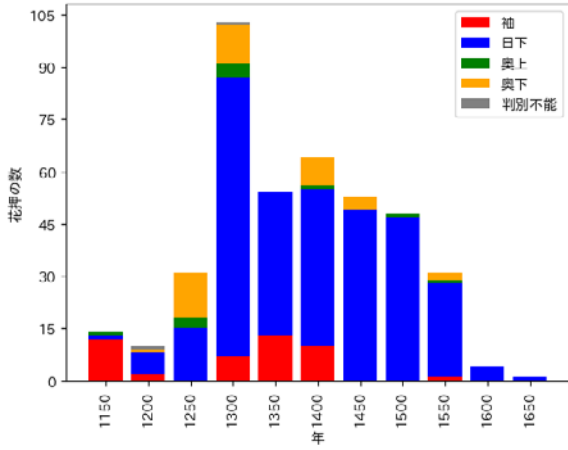


図8 花押数の時系列変化と位置の内訳  
Figure 8 Time-series changes in the number of signatures and a breakdown of their position.

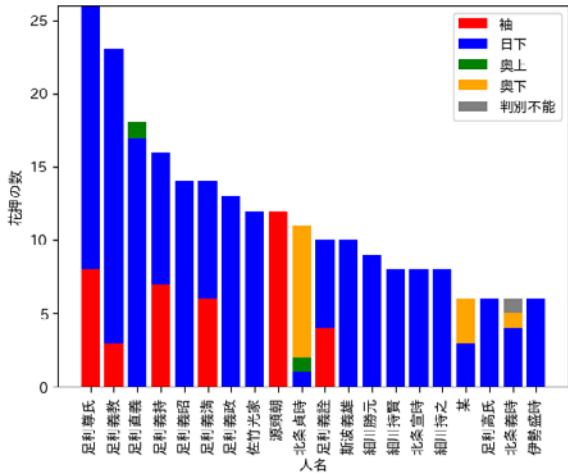


図9 人物別の花押数と位置の内訳  
Figure 9 Number of signatures by person and a breakdown of their position.

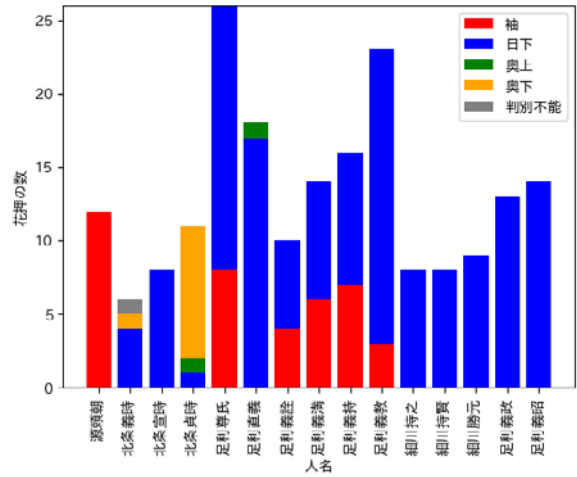


図10 人物別の花押数と位置の内訳 (人物の生年で並び替え)  
Figure 10 Number of signatures by person and a breakdown of their position (Sort by person's birth year).

さらに、1200-1300年および北条義時と貞時による奥下の花押が集中している点については、鎌倉幕府の第2代執権である北条義時(在職:1205-1224年)および第9代執権である貞時(在職:1284-1301年)などによる、将軍の意を受けて出した公的な文書である「関東御教書」などが該当するためである。関東御教書では、日下ではなく奥下に署名をすることが多く、この特徴が可視化されている。図11は北条貞時の花押数の時系列変化を可視化した結果である。北条貞時が執権を務めていた時期と奥下花押の対応が確認できる。

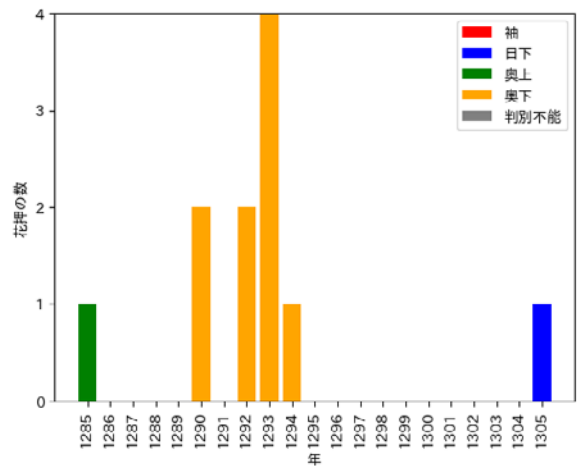


図11 北条貞時の花押数の時系列変化  
Figure 11 Time-series changes in the number of signatures by Hōjō Sadatoki.

### 4.3. 花押の大きさに関する分析

3章において、料紙の領域をデータ化したため、料紙に対応する花押の大きさの比率を算出することができる。図12に位置別の料紙に対する花押の面積比率の平均値を可視化した結果を示す。図中のバーは標準偏差を示す。

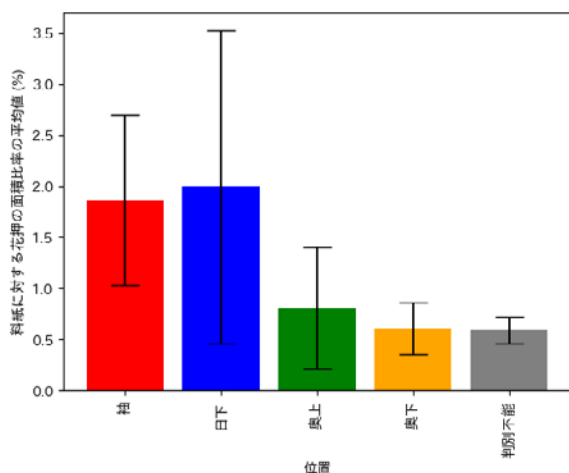


図12 料紙に対する花押の面積比率の平均値 (位置別)

Figure 12 Average ratio of area of signatures to paper (by position).

図12から、袖および日下のほうが、奥上および奥下の花押より大きく描かれることが多いことがわかる。また袖と日下の花押の大きさを比較すると、平均値に大きな違いはないが、日下の花押は大小にばらつきがあることがわかる。この理由として、日下花押は身分の高低に関わらずに使用され、一般的には小さく描かれるが、足利尊氏などは晩年に大きな花押を日下を書く、という特徴に基づくものと考えられる。今回は身分の高い人物からの文書が比較的多いため、このようなばらつきの違いが生じたと考えられる。

図13に人物(花押数の多い上位20名)別の料紙に対する花押が占める割合の比率の平均値を可視化した結果を示す。図12と同様、図中のバーは標準偏差を示す。

この結果、図9で袖の割合が高かった源頼朝と足利尊氏を比較すると、源頼朝の花押の大きさのばらつきは小さいが、足利尊氏のものはばらつきが大きいことがわかる。これは、前者は一貫して高い地位から文書を発給しているのに対し、後者は比較的低いところからはじまり、段階的に相対的地位が上がっていったことなどが理由として考えられる。これについて、図13の右から3番目に位置する「足利高氏」は尊氏の改名以前の名前であり、尊氏と高氏の大きさを比較すると、後者のほうが小さいことがわかる。今後、表記の違いを吸収した分析を行う必要もあるが、このよう

な傾向は、尊氏の人生と花押の大きさに連動があることを示す結果である。

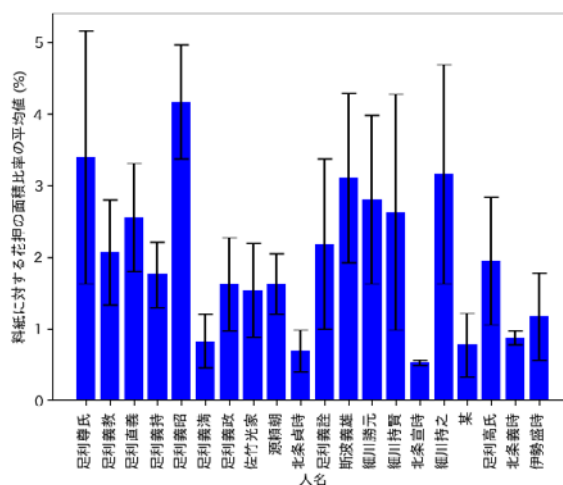


図13 料紙に対する花押の面積比率の平均値 (人物別)

Figure 13 Average ratio of area of signatures to paper (by person).

なお、図14に足利尊氏による花押の面積比率の時系列変化の結果を示す。将軍を務めた1338年から1358年のうち、特に晩年において、料紙に対する花押の面積比率が大きいことがわかる。

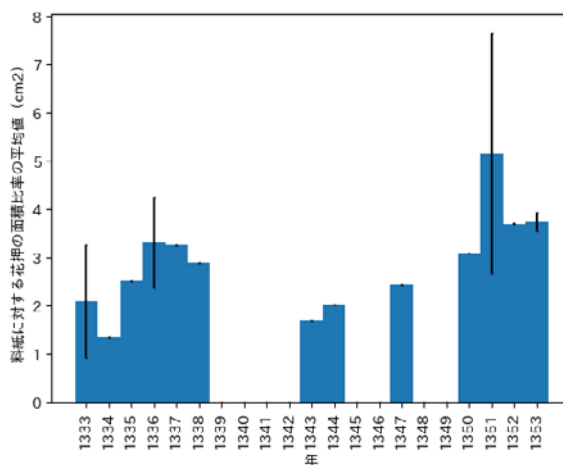


図14 料紙に対する花押の面積比率の平均値の時系列変化 (足利尊氏)

Figure 14 Time-series changes in the average ratio of area of signatures to paper (by Ashikaga Takauji).

また図13から、室町幕府第15代(最後の)将軍である足利義昭の花押の比率が比較的高いことがわかる。これは図15に示すように、義昭の

花押は点が右上に跳ね上がる特徴的な形状をしているため、矩形の面積が大きく算出される、という手法における課題の影響が含まれる。

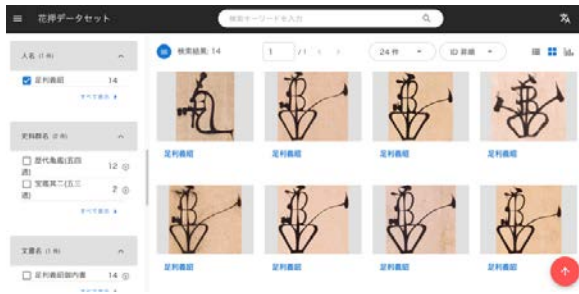


図 15 足利義昭の花押の例  
Figure 15 Example of signature of Ashikaga Yoshiaki.

さらに、室町時代の後期になり、料紙の大きさが小さくなっていくことに伴い、花押の占める比率が大きくなる傾向を示していると考えられる。この仮説の検証を目的として、画像中の定規の検出と読み取りを行い、料紙のおおよその実面積を算出した。その時系列変化を図 16 に示す。この結果から、1300 年から 1570 年頃にかけて、料紙の大きさの平均値が小さくなっていく傾向がみられる。しかし、足利義昭が将軍を務めた 1568 年から 1588 年ではその傾向が見られなかった。この点については、人物別の料紙の大きさの分析など、より詳細な分析を今後の課題とする。

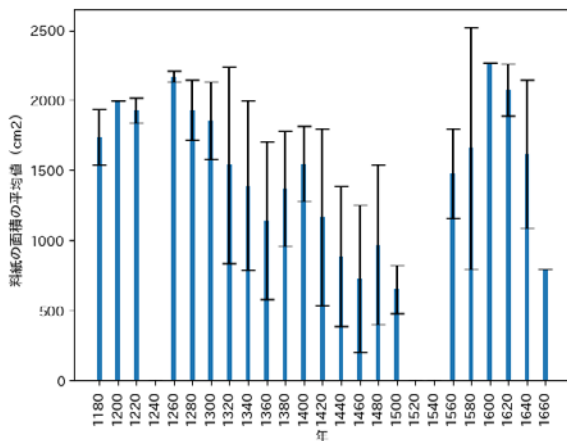


図 16 料紙の実面積の時系列変化  
Figure 16 Time-series changes in the actual area of the paper.

#### 4.4. 分析結果のまとめ

上述した各種分析について、総じて既存の研究成果・言説を支持する定量的な結果が得られた。これは花押に対する定量的な分析の有用性を支

持する結果であると考えられる。今後、より大規模なデータへの適用を通じて、新たな史実の発見につなげたい。

## 5. 発展

ここでは、これまでのデータ作成および分析結果に基づく発展的な取り組みについて取り上げる。

### 5.1. 物体検出モデルの開発

3 で作成したデータセットを用いて、史料画像中の料紙・花押・日付を検出するモデルを構築した。手法としては YOLOv5 を用い、図 17 に示す学習結果を得た。横軸はエポック数 (300) を示す。学習の結果、適合率 99%、再現率 99%、 $mAP@[0.5:0.95]$  97%を示すモデルが作成された。この結果から、過学習の可能性が懸念されるが、今後の利用を通じて検証を進めたい。

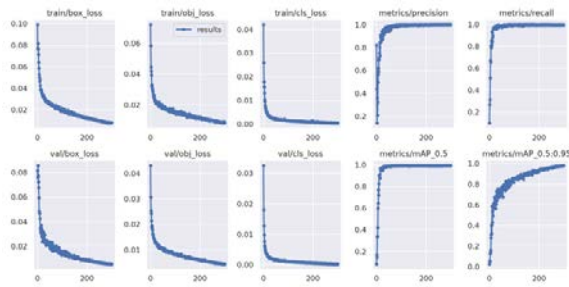


図 17 Yolo v5 の学習結果  
Figure 17 Training results of Yolo v5.

図 18 に同モデルを Hugging Face にデプロイしたアプリケーションによる推論の例を示す。料紙 (ryoshi\_main), 日付 (date) および花押 (kao) を検出できていることが確認できる。本研究で分析した要素を機械的に算出することにより、今後より大規模なデータ分析へと繋げる。また、本研究で作成したデータセットの公開も今後検討する。

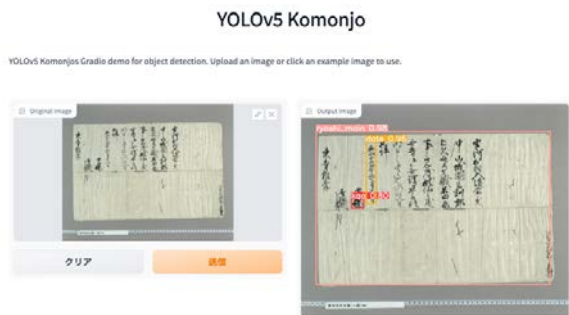


図 18 Yolo v5 による推論の例  
Figure 18 Example of inference with Yolo v5.

## 5.2. 料紙の実面積の自動算出に関する試み

図 16 において、料紙の実面積の時系列変化を試みた。本論文では紙面の都合上詳細を割愛するが、既存の研究[11]を参考に、史料画像に写る定規の検出および目盛の自動読み取りプログラムを開発している。前者は Yolo v5 を用いた物体検出、後者は古典的な画像処理技術による読み取りを行っている。今後、本手法に関する内容を別途報告するとともに、古文書に対する機械処理フローにおける 1 要素として位置づける予定である。

## 5.3. 検索アプリケーションの開発

これまで作成したデータを検索可能なアプリケーションを構築した。図 15 が本アプリケーションの画面例である。既存の花押データベースが提供する機能との差分として、本研究の成果に基づき、文書形式（縦紙・折紙）や花押の位置（袖・日下・奥上・奥下）に基づく絞り込み機能を提供する。また図 19 に示す、関連する花押の提示機能を提供する。特に画像の類似度に基づく関連アイテムの提示にあたっては、仲居らの手法[10]を使用して、VGG 16 と Annoy を用いて抽出した特徴ベクトルに基づく類似花押検索の仕組みを構築した。この仕組みを活用し、花押に対する人物の機械的な同定など、今後より大量の花押を対象とした定量的な分析につなげたい。

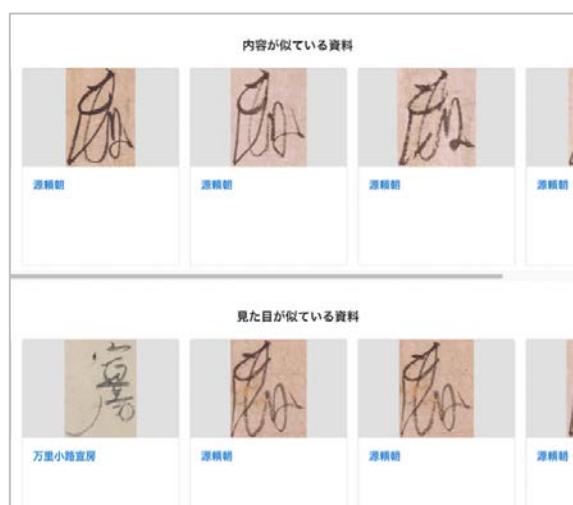


図 19 関連する花押の提示例  
Figure 19 Example of display of related signature.

## 6. 結論

本研究では、花押の検出や分類など、花押に対する定量的な分析を試みることで、データ駆動型の歴史情報研究のモデル検証を試みた。古文書に関する様式の一部をデータ化し、データの位置や大きさに基づく探索的データ分析を行うことで、

花押に関する既存の研究成果や言説を再現・支持する結果を得ることができた。また、さらなる大規模なデータ分析に向けた仕掛けとして、機械学習技術を用いた花押の類似画像検索および文書様式の検出モデルを作成した。

今後、本研究の手法および機械学習モデルを活用することで、花押に対する定量的な分析を行い、新たな史実の発見に寄与することを目指す。

## 謝辞

本研究は JSPS 科研費 19K20626, 21K18014 および東京大学史料編纂所「データ駆動型歴史情報研究基盤の構築」による成果の一部である。

## 参考文献

- [1]. データ駆動型歴史情報研究基盤の構築[東京大学]. [https://www.u-tokyo.ac.jp/adm/fsi/ja/projects/sdgs/project\\_00214.html](https://www.u-tokyo.ac.jp/adm/fsi/ja/projects/sdgs/project_00214.html), (参照 2022-08-30).
- [2]. 北本朝展, 鈴木親彦, 寺尾承子, 堀井美里, 堀井 洋, 地理的史料を対象とした歴史地名の構造化と統合に基づく江戸ビッグデータの構築, じんもんこん 2020 論文集, Vol.20 20, pp.171-178, 2020.
- [3]. 中村寛, 須田牧子, 黒嶋敏, 井上聡, 山田太造. データ駆動型歴史情報研究基盤の構築に向けた知識ベースの構築とその活用: 絵図史料を対象として, じんもんこん 2021 論文集, Vol.2021, pp.88-95, 2022.
- [4]. 佐藤進一. [新版] 古文書学入門, 法政大学出版社, 2003.
- [5]. 佐藤進一. 増補 花押を読む, 平凡社ライブラリー, 2020.
- [6]. 上島有. 中世花押の謎を解く一足利将軍家とその花押一, 山川出版社, 2004.
- [7]. 鬼塚洋輔, 大山航, 山田太造, 井上聡, 内田誠一. 花押類似検索のための畳み込みオートエンコーダによる画像特徴抽出. じんもんこん 2018 論文集, Vol.2018, pp.257-262, 2018.
- [8]. 岡野元春, 村本俊哉, 村川猛彦. 花押を対象とした画像比較システムの構築. じんもんこん 2021 論文集, Vol.2021, pp.212-217, 2021.
- [9]. nbdev, <https://nbdev.fast.ai/>, (参照 2022-10-24).
- [10]. 仲居優志, 宮崎肇, 井上聡, 大山航. 深層計量学習を用いた花押画像検索, 情報処理学会 第 84 回全国大会, 2022.
- [11]. 永崎研宣, 青池亨, 本間淳. デジタルアーカイブにおける実寸比較の実現に向けた実装と課題, 研究報告人文科学とコンピュータ (CH), Vol. 2022-CH-129, No.13, pp.1-3, 2022.