Implementation of data-driven historical informatics research on Kao (Stylized Signature)

Nakamura, Satoru
na.kamura.1263@gmail.com
The University of Tokyo, Japan

Liu, Guanwei
k-ryu@hi.u-tokyo.ac.jp
The University of Tokyo, Japan

Miyazaki, Hajime
h-miyazaki@g.ecc.u-tokyo.ac.jp
The University of Tokyo, Japan

Inoue, Satoshi
inoue@hi.u-tokyo.ac.jp
The University of Tokyo, Japan

Ohyama, Wataru
w.ohyama@mail.dendai.ac.jp
Tokyo Denki University

Yamada, Taizo
t_yamada@hi.u-tokyo.ac.jp
The University of Tokyo, Japan

Kao is a stylized signature used in East Asian cultures. Figure 1 shows examples of documents and signatures. The Historiographical Institute the University of Tokyo has the largest collection of data on these signatures in Japan. As of November 2022, data are available for approximately 26,000 signatures.

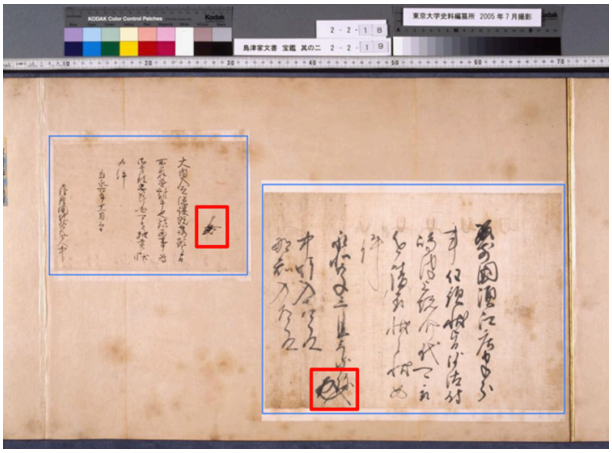

Figure 1. examples of documents and Kao (stylized signatures).

As shown in Figure 2, these signatures are classified based on location. For example, they are classified as "Sode", which refers to the right margin of the document, "Oku-ue," which refers to the top of the line following the date, "Oku-sita," which refers to the bottom of the line following the date, and "Nikka," which refers to the bottom of the date. Kao which is located on Sode, is the most dignified form for the addressee. On the other hand, documents with the "Oku-ue" style are more respectful to the addressee than those with the “Nikka" and "Oku-shita" styles.

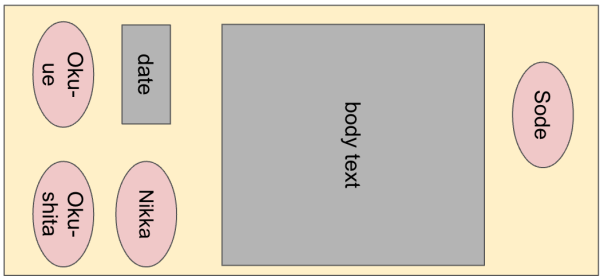

Figure 2. Classification by location of Kao.

It is a common understanding among paleography and Japanese historiography that the position and size of the Kao are essential elements of the manner of writing and important information indicating the high or low status of the issuers and their relationships with the recipients. However, one of the challenges is that, although many studies on Kao have been conducted [1], not enough data has been presented for such a common understanding.

Therefore, this study quantitatively analyzes the location and size of Kao. In this study, information about the location and size of the Kao in 412 historical material images is created, and data is explored using numerical summarization and visualization through exploratory data analysis to identify potential relationships among variables. We used nbdev [2] to summarize and visualize numerical values, as shown in Figure 3.

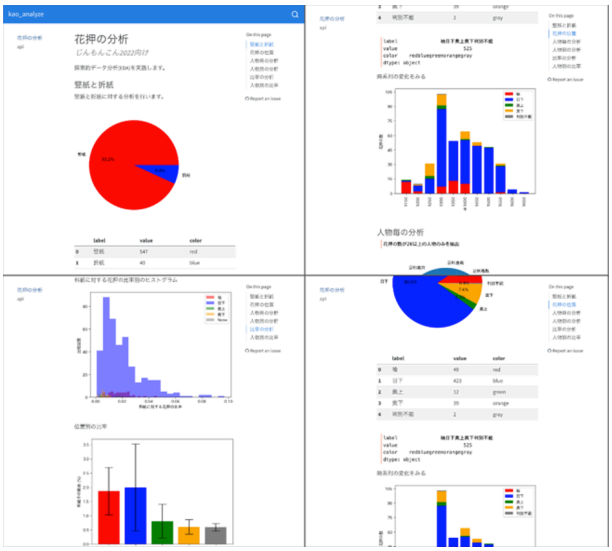

Figure 3. Example of exploratory data analysis.

This provided quantitative support for the existing research findings. For example, Figure 4 shows an average ratio and standard deviation of the area of signatures to paper by their position. The Sode and Nikka Kaos are often larger than the Oku-ue and Oku-shita ones. Comparing the sizes of Sode and Nikka Kao, there is no significant difference in the average value, but there is a large difference in the standard deviation. This may be partly due to the fact that the Sode Kao is used mainly by people of high status, whereas the Nikka Kao is used regardless of status. In addition, an analysis of the location and size of the Kao by a person has also been conducted.

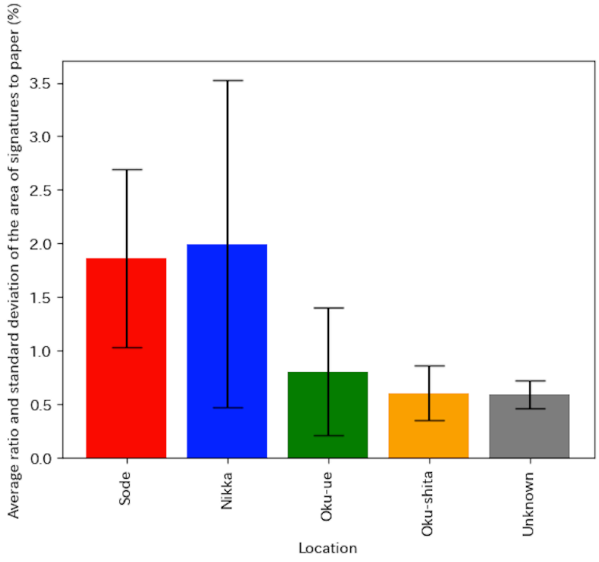

Figure 4. Average ratio and standard deviation of the area of signatures to paper.

We also attempted to build a machine-learning model to apply this method to a large amount of data. Specifically, we built a system for extracting Kao using an object detection model and a search system for similar Kao for personal identification of them.

Figure 5 shows an example of object detection using YOLO v5 [3].

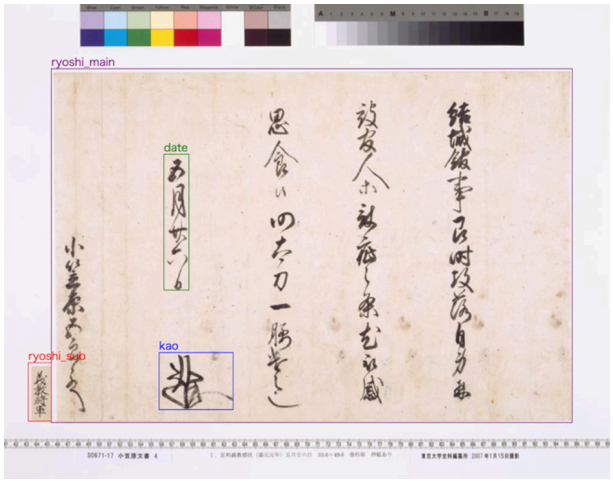

Figure 5. Example of object detection using YOLO v5.

In this presentation, we report on the results of the above analysis and discuss the potential of data-driven historical research.

Bibliography

[佐藤進一. ［新版］古文書学入門, 法政大学出版局, 2003. (in Japanese)

nbdev, https://nbdev.fast.ai/, Accessed on 2022-11-05.

Ultralytics, YOLOv5, https://github.com/ultralytics/yolov5, Accessed on 2022-11-05.