Humans, machines, and the classification of Celtic coinage – the ClaReNet project

Felix Riede, Sébastien Plutniak (ID) *and Shumon Tobias Hussain* (ID) *based on peer reviews by Eythan Levy* (ID) *and Alex Brandsen* (ID)

The paper entitled "IT- and machine learning-based methods of classification: The cooperative project ClaReNet – Classification and Representation for Networks" submitted by Chrisowalandis Deligio and colleagues presents the joint efforts of numismatists and data scientists in classifying a large corpus of Celtic coinage. Under the project banner "ClaReNet – Classification and Representation for Networks", they seek to explore the possibilities and also the limits of new computational methods for classification and representation. This approach seeks to rigorously keep humans in the loop in that insights from Science and Technology Studies inform the way in which knowledge generation is monitored as part of this project. The paper was first developed for a special conference session convened at the EAA annual meeting in 2021 and is intended for an edited volume on the topic of typology, taxonomy, classification theory, and computational approaches in archaeology.

Deligio et al. (2024) begin with a brief and pertinent research-historical outline of Celtic coin studies with a specific focus on issues of classification. They raise interesting points about how variable degrees of standardisation in manufacture – industrial versus craft production, for instance – impact our ability to derive tidy typologies. The successes and failures of particular classificatory procedures and protocols can therefore help inform on the technological contexts of various object worlds and the particularities of human practices linked to their creation. These insights and discussion are eminently case-transferrable and applies not only to coinage, which after all is rather standardised, but to practically all material culture not made by machines – a point that gels particularly well with the contribution by Lucas (2022) to be published in the same volume. Although Deligio et al. do not flag this up specifically, the very expectation that archaeological finds should

neatly fall into typological categories likely related to the objects that were initially used by pioneers of the method such as Montelius (1903) to elaborate its basic principles: these were metal objects often produced in moulds and very likely by only a very small number of highly-trained craftspeople such that the production of these objects approached the standardisation seen in industrial times (Nørgaard 2018) – see also Riede's (2023) contribution due to be published in the same volume.

At the core of Deligio and colleagues' contribution is the exploration of machine-aided classification that should, via automation, assist in handling the often large numbers of coins available in collections and, at times, from single sites or hoards. Specifically, the authors discuss the treatment of the remarkable coin hoard from Le Câtillon II on Jersey consisting of approximately 70,000 individual Celtic coins. They proceeded to employ several advanced supervised and unsupervised classification methods to process this stupendously large number of objects. Their contribution does not stop there, however, but also seeks to articulate the discussion of machine-aided classification with more theoretically informed perspectives on knowledge production. In line with similar approaches developed in-house at the Römisch-Germanische Kommission (Hofmann 2018; Hofmann et al. 2019; Auth et al. 2023), the authors also report on their cutely acronymed PANDA (Path Dependencies, Actor-Network and Digital Agency) methodology, which they deploy to reflect on the various actors and actants – humans, software, and hardware – that come together in the creation of this new body of knowledge. This latter impetus of the paper thus lifts the lid on the many intricate, idiosyncratic, and often quirky decisions and processes that characterise research in general and research that brings together humans and machines in particular – a concrete example of the messiness of knowledge production that commonly remains hidden behind the face of the published book or paper but which science studies have long pointed at as vital components of the scientific process itself (Latour and Woolgar 1979; Galison 1997; Shapin and Schaffer 2011). In this manner, the present contribution serves as inspiration to the many similar projects that are emerging right now in demonstrating just how vital a due integration of theory, epistemology, and method is as scholars are forging their path into a future where few if any archaeological projects do not include some element of machine-assistance.

References:

Auth, Frederic, Katja Rösler, Wenke Domscheit, and Kerstin P. Hofmann. 2023. "From Paper to Byte: A Workshop Report on the Digital Transformation of Two Thing Editions." Zenodo. https://doi.org/10.5281/zenodo.8214563

Deligio, Chrisowalandis, Caroline von Nicolai, Markus Möller, Katja Rösler, Julia Tietz, Robin Krause, Kerstin P. Hofmann, Karsten Tolle, and David Wigg-Wolf. 2024. "IT- and Machine Learning-Based Methods of Classification: The Cooperative Project ClaReNet – Classification and Representation for Networks." Zenodo, ver.5 peer-reviewed and recommended by PCI Archaeology https://doi.org/10.5281/zenodo.7341342

Galison, Peter Louis. 1997. Image and Logic: A Material Culture of Microphysics. Chicago, IL: University of Chicago Press.

Hofmann, Kerstin P. 2018. "Dingidentitäten Und Objekttransformationen. Einige Überlegungen Zur Edition von Archäologischen Funden." In Objektepistemologien. Zum Verhältnis von Dingen Und Wissen, edited by Markus Hilgert, Kerstin P. Hofmann, and Henrike Simon, 179–215. Berlin Studies of the Ancient World 59. Berlin: Edition Topoi. https://dx.doi.org/10.17171/3-59

Hofmann, Kerstin P., Susanne Grunwald, Franziska Lang, Ulrike Peter, Katja Rösler, Louise Rokohl, Stefan Schreiber, Karsten Tolle, and David Wigg-Wolf. 2019. "Ding-Editionen. Vom Archäologischen (Be-)Fund Übers Corpus Ins Netz." E-Forschungsberichte des DAI 2019/2. E-Forschungsberichte Des DAI. Berlin: Deutsches Archäologisches Institut. https://publications.dainst.org/journals/efb/2236/6674

Latour, Bruno, and Steve Woolgar. 1979. Laboratory Life: The Social Construction of Scientific Facts. Laboratory Life: The Social Construction of Scientific Facts. Beverly Hills, CA: Sage.

Lucas, Gavin. 2022. "Archaeology, Typology and Machine Epistemology." Zenodo. https://doi.org/10.5281/zenodo.7622162

Montelius, Gustaf Oscar Augustin. 1903. Die Typologische Methode. Stockholm: Almqvist and Wicksell.

Nørgaard, Heide Wrobel. 2018. Bronze Age Metalwork: Techniques and Traditions in the Nordic Bronze Age 1500-1100 BC. Oxford: Archaeopress Archaeology.

Riede, Felix. 2023. "The Role of Heritage Databases in Typological Reification: A Case Study from the Final Palaeolithic of Southern Scandinavia." Zenodo. https://doi.org/10.5281/zenodo.8372671

Shapin, Steven, and Simon Schaffer. 2011. Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life. Princeton: Princeton University Press.

Reviews

Evaluation round #1

Authors' reply, 12 August 2024

We thank the reviewers for their constructive comments, which have helped to significantly improve and structure the manuscript. We have taken most of them on board, in particular drastically shortening the section on PANDA, the Science and Technology Study. We have also explained that while similar work involving image recognition and numismatics has been done, it has never been specifically targetted to solve a problem in the way our project did. Nor has anyone ever worked with this kind of material.

One of the reviewers wanted more information on some of the processes involved, for example how we arrived a a particular value for K. We discussed this together and all felt that our paper is very much a report on the first stages of the project, when such things were often a case of trail and error, and were later optimised. Discussion at length would have meant writing a very new paper, integrating later results.
Such matters will be explained and discussed at length in the extensive final report on the project that we are preparing at present (the project ended in January).

The question was also raised as to whether we address a computer or numismatic readership. The answer is to some extent both! For example, a reader coming from the computer angle will not have the numismatic background and thus needs explanations of what a die-study is. At the same time, we would hope that the contribution is written in such a way that numismatists will grasp enough to understand the enormous potential that such methods have in the field.

If you need any more specific answers we are happy to go into more detail.

Decision by Felix Riede, Sébastien Plutniak and Shumon Tobias Hussain, posted 20 March 2023, validated 20 March 2023

Please revise

Dear Dr. Deligio, dear author team,

two very favourable reviews of your contribution have now ticked in, each f which makes constructive and useful suggestions as to how to further improve your manuscript. One reviewer added comments in-text, which I also append here.

We look forward to seeing your revised version in due time.

Best wishes,

Felix **Download recommender's annotations**

Reviewed by Alex Brandsen ⓘ, 16 March 2023

I thoroughly enjoyed reading this paper, it is in interesting approach in trying to help humans with classification without replacing them completely, which is the "human-in-the-loop" type machine learning that could really benefit archaeologists. The methods and results presented are preliminary, but promising: the authors apply clustering and supervised learning to image data of coins, and find that these methods seem to cluster relatively well. The introduction sufficiently introduces the problem and background, and the STS study is an interesting addition.

I do have some thoughts on improvements that would further strengthen the paper:

"AI" is used quite often, but might be too vague. I understand you want to mention it, and that's ok in the abstract / start of introduction, but use the names of the actual techniques you use in rest. i.e. p3 line 19: "the AI" should be k-means clustering, or just unsupervised learning, or even machine learning.

Depending on what journal this will be submitted to, perhaps the numismatic background can be dialled back, or the explanations of e.g. the CNNs can be reduced. Currently, the CNN / k-means section is not easily understood by an average archaeologist, so this should be summarised and/or simplified if submitting to an archaeology journal (as opposed to a more computer science focused journal). For example, you explain a dense layer, but to do so use terms like 'fully meshed', 'weights' and 'neuron', which someone without a computer science background will not know, but for a CS journal, this is too much general knowledge and shouldn't be included in the paper. It would be worth tailoring the background sections to the journal you plan to submit it to.

p16, line 7: you choose an arbitrary k, but since you have ground truth labels, the optimal k can be calculated, see e.g. `https://aiaspirant.com/optimal-k-in-k-menas/` (unless the CNN part of DeepCluster completely invalidates these methods for some reason, then please mention that in the text)

p16, line 12: "also intended to explore the possibilities of employing an unsupervised method", but clustering is already an unsupervised method? Or did you mean a supervised method?

p16/17: you provide some metrics on the overlap between clusters and labels, but these are incomplete (not all clusters are described), and you don't use standard metrics, which makes it impossible to compare this study to similar studies, or to evaluate the effectiveness of your methods as a reader. I'm not an expert on evaluating clusters with labels, but I believe the adjusted Rand score is often used. It would be good to calculate this for the whole dataset, but also per class, as some classes might be more difficult to cluster. (this is assuming you have a label for each coin, if not, disregard this comment) The same goes for the die study supervised experiment.

Fig. 15: the axes are unlabelled, which made it more difficult to understand the graphs. Also from the describing text it wasn't clear that these graphs show the count of labels in different clusters, I initially expected some sort of accuracy metric. Please add axis labels and clarify the surrounding text.

As far as I can see, the code (and data?) are not available, or at least not directly referenced from the paper. Please add this, in the context of open, reproducible science.

The PANDA study was interesting, as I hadn't heard of this before. However, to me it was a bit unexpected, as it is (in some ways) separated from the end goal of the classification of coins. This could be a separate paper, so the focus of this paper is clearer, but I'll leave that choice up to the authors.

Instead, an analysis of how helpful the classification/clustering is for helping the numismatics classify the whole hoard would be more interesting and would strengthen the paper (although I also understand this is preliminary, and that might be done in a later paper). Things I'd like to see is how they interact with the computer/classifications, how much time was saved, whether or not pre-existing classification schemes and cognitive skills are challenged by the machine learning classifications, etc.

Some minor typos/corrections:
- 'k' is not always italic
- p17, line 23: "Furthemore" should be "Furthermore"

Reviewed by **Eythan Levy** [iD], 08 March 2023

Very interesting and relevant paper. I have included all my comments directly in the attached PDF.

My main concerns are:

1. I think that the results of the two AI experiments need to be discussed more in-depth.

2. I have reservations about the PANDA section, which I have found less interesting than the AI section, and which seemed a bit off-topic to me. But see my comments in the PDF for full details.

Download the review