The state of the art: Evaluating Machine Learning approaches to lithic sourcing in the UK

Tom Elliot  
(tom.elliot@outlook.com)  
Robert Morse  
(rob.632@hotmail.co.uk)
Structure

The Mesolithic, PhD work and results

Dimensionality reduction

Novelty detection model

Evaluation of three Machine Learning classifiers

Conclusions and future directions
Welcome to the Mesolithic

- ‘Meso’ - Middle - ‘Lithic’ - Stone = Middle Stone Age
- Around 9600 - 3800 BC
- Modern Human (Homo Sapiens Sapiens)
- Nomadic Hunter-Gatherers
- ‘Band’ size communities - fifteen to fifty people
Ethnography
Reconstructions
Cheddar Man - Gough’s Cave
Climate Change

TEMPERATURE CURVE LAST FIFTEEN THOUSAND YEARS

- Present temperature
- Holocene Warm Period
- Roman Warm Period
- Medieval Warm Period
- Little Ice Age
- Present global warming

Mesolithic
Environment
The Mesolithic in the Marches

- PhD at the University of Worcester
  - Aimed to identify the sources of Mesolithic flint artefacts from sites in the Lower Wye Valley region to evaluate previous hypotheses of mobility.

- Compared artefacts with geological samples to identify their probable geological source.

- Used Laser Ablation Inductively-Coupled Plasma Mass Spectrometry (LA-ICP-MS) and the Machine Learning technique Random Forest (Breiman 2001)
The problem - Where were the artefacts sourced from?
Previous hypotheses of flint sources used
Artefact Sample Sites

- 12 Archaeological Sites
- 118 artefacts in total
- Microliths or artefacts from secure Mesolithic contexts
- Includes surface finds due to the limited number of excavated sites in the Study Area
- Neolithic artefacts from Wellington Quarry used as a comparison
Superficial Geology Sample Sites

- 14 Superficial Geology sites
- 145 nodules of flint
- River terraces (BGS Britannia Catchments Group (River Severn and tributaries))
- Glacial deposits of the Albion Glaciogenic group
- Beach samples from near Nab Head, Pembrokeshire (thanks to Andrew David)
Bedrock Geological Sample Sites

- 21 Chalk Bedrock Geology sites
- 269 nodules of flint
- Focussed on the South-West, following previous hypotheses
- Mostly SSSI Sites - linking to stratigraphic logs and previous geological research
Bedrock Sampling

Sampling at Peacehaven near Brighton

Sampling 114m up, at Beer Head in Devon
Laser Ablation Inductively-Coupled Mass Spectrometry

- Analysis in triplicate
  - 3 measurements per nodule or artefact
  - 40-60 seconds dwell time, 120 washout
  - 1597 measurements in total

- 53 Elements (used as features in ML)

- NewWave NWR213 Laser Instrument
- Perkin Elmer NexION 300Q Mass Spec
- Data reduction in GLITTER and LADR v0.6
Random Forest Analysis

- The Random Forest methodology achieved 0.68 F1-score in correctly identifying the geological sample site for each individual measurement.

- Importantly, this included ruling out other locations.

- 95% of the time, at least two of the three measurements for the artefacts agreed on the source.

- Artefact measurements were then averaged to provide the final source determination for the PhD.
PhD Results

- Short, medium and long-distance (200km+) movement of raw materials throughout Mesolithic and into the Neolithic.
- Materials primarily from river terrace deposits, though the geology represented by Peacehaven features heavily too.
- Movement of materials well beyond all previous hypotheses of mobility for the region.
- Archaeologically, this fits with long distance movements of materials across Europe.
- Based on global trends in ethnographic research (Topping 2017), long-distance movement likely represents exchange practices and a more complex picture than previously considered.

*** PhD results superseded by those presented in this research (in prep!) ***
Evaluating more techniques

- We’ve since improved on the methodology and results of the PhD

- Here we present a robust evaluation of three popular Machine Learning techniques, using the geological data from the PhD to see which technique performs the best.
  - Random Forest (RF)
  - K-Nearest Neighbour (KNN)
  - Support Vector Machines (SVM)
definitions

- **Classes:** geological sources that classification algorithm can classify into

- **F1 score:** evaluation metric for assessing classifier performance, closer to 1 or (100%) is better

- **Observation:** a row of training data

- **Feature:** a column that is predictive of where the sample is from, in this case mass spectrometry data
We used t-Distributed Stochastic Neighbour Embedding (t-SNE) to first visualise the data. This demonstrated that the bedrock data showed good structure, while the superficial deposit data structure was less well resolved. Only the bedrock data was therefore used to evaluate the ML techniques and train the final model.

**Figure 1**

A. t-SNE Analysis of Bedrock Sample Data

B. t-SNE of Superficial Sample Data
Feature selection - How many columns?

- From the Mass Spectrometry data for each sample, we used to Recursive Feature Elimination (RFE) to establish which elements (Iron, Aluminium etc.) are the most useful in differentiating between the geological sites.

- Gini impurity was used as the feature importance

- This shows 15 elements which are primarily useful.
These match most of the elements identified as important by Rockman (Rockman 2003) in Britain and Brandl (Brandl et al. 2018) in Europe. This somewhat validates the feature importances.

The ability to assign feature importances is a large advantage of the random forest method.

We used the top 15 features for the all models from this point onwards.

<table>
<thead>
<tr>
<th>Features (Elements)</th>
<th>RFE Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zr90, Ba137, Sr88, Ge72, Cr52, S33, U238, Al27, B11, Mg24</td>
<td>1</td>
</tr>
<tr>
<td>Nd146</td>
<td>2</td>
</tr>
<tr>
<td>Sc45</td>
<td>3</td>
</tr>
<tr>
<td>K39</td>
<td>4</td>
</tr>
<tr>
<td>Pr141</td>
<td>5</td>
</tr>
<tr>
<td>Li7</td>
<td>6</td>
</tr>
<tr>
<td>V51</td>
<td>7</td>
</tr>
<tr>
<td>P31</td>
<td>8</td>
</tr>
<tr>
<td>Mn55</td>
<td>9</td>
</tr>
<tr>
<td>Cd111</td>
<td>10</td>
</tr>
<tr>
<td>La139</td>
<td>11</td>
</tr>
</tbody>
</table>
Difference from Kaggle

- In Kaggle competitions it is usually observations that can be classified correctly into one of the classes represented in the training data.

- For example in the titanic binary classification problem, passengers have to be classified as either dead or alive.

- It is different for our problem, a proportion of artefacts may have originally been sourced from a deposit that we had not sampled from

- We solved this problem by implementing a novelty detection model called Local Outlier Factor.
Local outlier factor model

- Train model on training data aka geological sample data.
- Predict on artefact data, classes are outlier or not outlier
- Takes k nearest neighbors for each artefact.
- Compares density of artefacts with that of local density of training data to give an outlier score
- Outliers were classified as ‘other’ source

Novelty (outlier) detection

- We used novelty detection to detect samples that do not belong to the geological sites, potentially demonstrating materials from other sources.

- Importantly, to our knowledge, this has not been used in previous studies of lithic sourcing.
To assess whether collecting more samples would help improve model performance you can look at the learning curve.

Increase size of training data, evaluate model on train and test data.
Assessing the value of collecting more samples

- Class abundance (number of sample measurements per geological site) versus the class F1-score (how well the model correctly identified the right geological site).
- This shows a positive correlation.
- More samples = better classification performance for most classes (sample site)
Final Model Evaluations

- 10-fold stratified cross-validation
- 90/10 training/testing data split
- Random Forest was the best performing technique with an F1 score of 0.86

K-fold Cross validation (Gufosowa, 2019, wikipedia commons)
Model evaluation overview

Geo
696
CV 10-fold

Art
363
model evaluation + prediction

hyperparameter optimisation

CV 5-fold
for hyperparameter optimisation

10-fold optimisation
Model Evaluations

A. RFC

B. SVM

C. KNN

Class Specific/Individual Sample Site Evaluations

D.

E.

F.

Geological Site

Geological Site

Geological Site
Conclusions

- Random Forest was the best machine learning technique
- More geological sampling will help.
  - Greater sampling of under-represented sites
- Flint in Britain can be sourced down to specific geological facies.
  - Our data was only grouped by geological site as more samples are needed
- Caution still needed towards data and analysis
  - Patina/ patination, geochemical and taphonomic effects
  - Garbage in = garbage out
  - Biases
Future directions

- Adoption of standardised and robust evaluation methodologies so different machine learning techniques can be compared
- Development of an open-source data-store, allowing for third-party collaboration
- Website and database, visualisations
Website -

- Built with flask, visualisations with plotly
- Present the data and results from the PhD (Elliot 2019) and ongoing research (Elliot, Morse and Smythe, forthcoming).
- Visualisations of the data (t-SNE plots, PCA).
- Uploading of third party artefact data, enable sourcing determinations to our geological data using our Random Forest model.
- A database of artefact sourcing determinations that is updated as new geological samples/data and new statistical approaches are added and developed (akin to how Radiocarbon dating models are updated).
Welcome!

This is a web app that allows visualisation of archeological research data.

This website provides an interface to enable interactive visualisation of mass spectrometry data collected for archeological research purposes. In order to jump straight into visualising the data head over to the visualisations page. If you would like to learn more about what the visualisation is showing and where and why the data was collected then head over to the Visualisation Details and Research pages.
Acknowledgements

Duane Smythe, Ashley Norris, Phil Holdship, Bernie Wood, University of Oxford
Department of Earth Sciences
Jodie Lewis, University of Worcester
Chantal Conneller, University of Newcastle
Nick Barton, University of Oxford
Elizabeth Walker, National Museum of Wales
Oliver Blackmore, Newport Museum and Art Gallery
Judy Stephenson, Hereford Museum Service
Robin Jackson, Worcester Historic Environment and Archaeology Service
Gerry Ronan
Roger Pye
Alex Henderson, Simon Bailey, Julie Birchenall, Caroline Rosen
Funding

- Lithic Studies Society - John Wymer Bursary
- British Cave Research Association - Cave Science and Technology Research Initiative (CSTRI)
- Robert Kiln Charitable Trust
- University of Worcester
Reference List