Chemometrics and Intelligent Laboratory Systems 146 (2015) 418–425

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

Software Description

An online application for the classification and evidence evaluation of forensic glass fragments

Gary Napier [a,*], Agostino Nobile [b], Tereza Neocleous [a]

[a] School of Mathematics and Statistics, University Gardens, University of Glasgow, Glasgow G12 8QW, UK
[b] Department of Mathematics, University of York, Heslington, York YO10 5DD, UK

ARTICLE INFO

Article history:
Received 11 April 2015
Received in revised form 12 June 2015
Accepted 18 June 2015
Available online 23 June 2015

Keywords:
Bayes factor
Chemical composition
SEM-EDX analysis

ABSTRACT

We present an easy-to-use and freely accessible online application for the analysis of forensic glass fragments. The application is browser based and takes as input .csv or .txt files containing measurements from glass fragments obtained using a scanning electron microscope with an energy-dispersive X-ray (SEM-EDX) spectrometer. The application was developed to (i) classify glass fragments into use-type categories (classification), and (ii) compute the evidential strength of two sets of fragments under competing propositions (evidence evaluation). Detailed examples of how to use the application for both tasks are given, which highlight its user-friendly interface. The suitability of the statistical methods used by the application was checked using simulation studies, and improvements upon previous methods were found in both tasks.

1. Introduction

Glass fragments are one of the many sources of forensic evidence. Fragments from a broken item can be recovered from a crime scene and sent to a forensic laboratory, where various measurements are recorded for analysis. Analysis of glass fragments is mainly focused on evidence evaluation, which involves computing the evidential strength of two sets of fragments (from the crime scene and from a suspect) under two competing propositions (the prosecution and defence propositions). Measurements obtained from glass fragments can also be used to help determine their use-type, thus providing additional information about the type of the glass item from which the fragments obtained from a suspect may have originated. As most glass fragments analysed are very small, their use-type cannot always be determined by their thickness or colour [1], and so measurements of physicochemical features are obtained. Here, focus is placed upon chemical composition measurements acquired from using a scanning electron microscope with an energy-dispersive X-ray (SEM-EDX) spectrometer [1]. The elemental composition data consist of the percentage weights (wt.%) of the main elements comprising a glass fragment.

In this paper, we present an easy-to-use online application for the purposes of (i) classifying glass fragments into use-type categories (classification), and (ii) computing the evidential strength of two sets of fragments under complementary propositions (evidence evaluation). The application is easily accessible and straightforward to use for both tasks. It is available at http://gnapier.shinyapps.io/GlassClassificationAndEvaluation/ and was developed using the shiny package, which is part of the statistical programming language R [2].

The paper is organised as follows: Section 2 describes the database used in the development of the application. Section 3 summarises the statistical model, and the classification and evidence evaluation methods developed in [3] and used by the application. Section 4 provides examples of how to use the application in the classification and evidence evaluation tasks. Section 5 discusses how the evidence evaluation results are reported. Concluding remarks are provided in Section 6.

2. Training data

The database used in the development of the application was provided by the Institute of Forensic Research, Krakow, and it consists of measurements obtained in an experimental setting using a scanning electron microscope with an energy-dispersive X-ray (SEM-EDX) spectrometer [1]. SEM-EDX analysis produces measurements, in the form of percentage weights (wt.%), on the main chemical elements that comprise the composition of a glass fragment. These are oxygen (O), sodium (Na), magnesium (Mg), aluminium (Al), silicon (Si), potassium (K), calcium (Ca) and iron (Fe). The database consists of glass fragments from 320 glass items across five use-types (26 bulbs, 94 car windows, 16 headlamps, 79 containers and 105 building windows). The chemical compositions of four glass fragments from each item were measured three times. Thus, the database has a hierarchical structure: three replicate measurements on four fragments from each of 320 glass items of five possible use-types.

* Corresponding author at: School of Mathematics and Statistics, University Gardens, University of Glasgow, Glasgow G12 8QW, UK.
E-mail address: gary.napier@glasgow.ac.uk (G. Napier).

http://dx.doi.org/10.1016/j.chemolab.2015.06.013
0169-7439/

As the measurements obtained from SEM-EDX are compositional, there is, as frequently happens with compositional data, a large number of zero measurements. Prior to model building, the percentage weights were transformed, by taking square roots of the ratios between each element weight and the weight of oxygen. The square root transformation was employed because it turned out to be more effective at stabilising the variability of the ratios; see [3] for further details on the choice of this transformation. The statistical model used by the application, and the methods used for classification and evidence evaluation, will be described in Section 3.

3. Methods

This section only provides a summary of the model and methods developed in [3], to which the reader is referred for full details.

3.1. Statistical model

A Bayesian mixed-effects model was used to account for the hierarchical structure of the database. The model incorporates a fixed effect for the mean of each use-type and three random effects: at item level, fragment level, and replicate measurement level. Denote the square root ratios from the k-th replicate measurement on the j-th fragment of the i-th glass item of use-type t by the p-dimensional vector $\mathbf{z}_{tijk}$. It is then assumed that

$$\begin{aligned}\mathbf{z}_{tijk} &= \boldsymbol{\theta}_t + \mathbf{b}_{ti} + \mathbf{c}_{tij} + \boldsymbol{\varepsilon}_{tijk},\\ \mathbf{b}_{ti} &\overset{\text{iid}}{\sim} N_p\left(\mathbf{0}, \Omega_t^{-1}\right),\ \mathbf{c}_{tij} \overset{\text{iid}}{\sim} N_p\left(\mathbf{0}, \Psi^{-1}\right),\ \boldsymbol{\varepsilon}_{tijk} \overset{\text{iid}}{\sim} N_p\left(\mathbf{0}, \Lambda^{-1}\right).\end{aligned} \tag{1}$$

The fixed effect for the mean of use-type t is denoted by $\boldsymbol{\theta}_t$; the item level random effect by $\mathbf{b}_{ti}$; the fragment within-item random effect by $\mathbf{c}_{tij}$; and the error at measurement level by $\boldsymbol{\varepsilon}_{tijk}$. The random effects are assumed to have multivariate normal distributions, with unknown precision (i.e. inverse covariance) matrices Ω_t, Ψ and Λ. Then, for a glass item $\mathbf{z}$ of use type $\mathcal{T}_{\mathbf{z}} = t$ with JK measurements, the distribution of $\mathbf{z}$ is

$$\mathbf{z}|\mathcal{T}_{\mathbf{z}} = t, \xi \sim N_{JKp}(\mathbf{1}_{JK} \otimes \boldsymbol{\theta}_t, \Sigma_t), \tag{2}$$

where $\xi = \{\boldsymbol{\theta}, \Omega, \Psi, \Lambda\}$ collectively denotes the model parameters and $\mathbf{1}_d$ is a column vector of d 1's. The covariance matrix Σ_t is given by

$$\Sigma_t = \left(\mathbf{1}_{JK}\mathbf{1}'_{JK}\right) \otimes \Omega_t^{-1} + \left[\mathbb{I}_J \otimes \left(\mathbf{1}_K \mathbf{1}'_K\right)\right] \otimes \Psi^{-1} + \mathbb{I}_{JK} \otimes \Lambda^{-1}, \tag{3}$$

where $\mathbb{I}_d$ is the $d \times d$ identity matrix.

The prior distributions placed on the fixed effects $\boldsymbol{\theta}_t$ are multivariate normal truncated to the positive orthant to ensure that the square root transformed means are non-negative:

$$\boldsymbol{\theta}_t \overset{\text{iid}}{\sim} N_p\left(\mathbf{0}, \Phi^{-1}\right), \quad \boldsymbol{\theta}_t > \mathbf{0}, \quad t = 1, \ldots, T.$$

The covariance matrix Φ^{-1} is fixed. The precision matrices for the random effects have conjugate Wishart priors placed upon them:

$$\Omega_t \sim W_p(d_{1t}, A_t), \quad \Psi \sim W_p(d_2, B), \quad \Lambda \sim W_p(d_3, C).$$

For more details on the prior and the Markov Chain Monte Carlo (MCMC) methods used see [3]. It is worth highlighting that the application does not need to run any MCMC as it uses the posterior draws obtained from modelling the database, thus making the application quick to use. The flip side of this point is that the application is not designed to re-estimate the model using a different background database, possibly available to the potential user.

As briefly mentioned in Section 2, the database contains a large proportion of zeros. To handle these zeros the background database was partitioned into subsets based on elemental configurations. The elemental configurations denote whether an element is present (above detection limit) or absent (below detection limit) from the composition of a glass item. The background database consists of glass items with ten different elemental configurations, as shown in Fig. 1. However, as the elements iron and potassium are responsible for the majority of the zeros, focus is placed on the presence or absence of these two elements only, thus reducing the number of configurations from ten to four. A Bayesian hierarchical model like (1) is then estimated for each subset of the background database for the four elemental configurations. For details on how the Bayesian hierarchical models for the four elemental configurations are brought together to form a composite model see [3].

3.2. Classification

Being able to predict the use-type of a glass fragment can help at the investigation stage of a legal case. To classify fragments, the application uses the posterior predictive distribution of the use type $\mathcal{T}_{\mathbf{y}}$ of a newly observed glass item's measurement vector $\mathbf{y}$ to be classified, conditional on the background database D described in Section 2, and the new item $\mathbf{y}$. Let $\mathbf{y}$ be a vector consisting of $\tilde{K}$ replicate measurements on each of $\tilde{J}$ fragments from the same glass item. The use-type probability of $\mathbf{y}$ is given by

$$\begin{aligned}p(\mathcal{T}_{\mathbf{y}} = t|\mathbf{y}, D) \propto p(\mathcal{T}_{\mathbf{y}} = t) &\frac{\alpha_{tm} + N_{tm}}{\sum_{r=1}^{M}(\alpha_{tr} + N_{tr})}\\ &\times E_{\xi_m|D_m}\left[p(\mathbf{y}|\mathcal{T}_{\mathbf{y}} = t,\ \mathcal{C}_{\mathbf{y}} = m, \xi_m)\right].\end{aligned} \tag{4}$$

The first two expressions on the right-hand side of (4) derive from modelling the counts N_{tm} of items in D that are of use-type t and configuration m. The expressions give the use-type probabilities for a newly observed glass item of use-type $\mathcal{T}_{\mathbf{y}}$ given it has elemental configuration $\mathcal{C}_{\mathbf{y}} = m$, without conditioning on the actual measurements $\mathbf{y}$. These use-

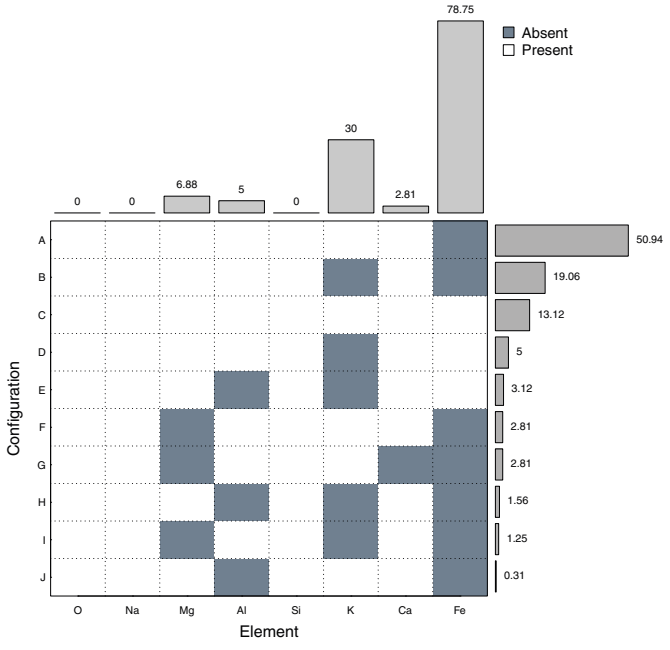

Fig. 1. Plot displaying the ten elemental configurations at item level observed in the glass database. The percentage of compositional zeros by element is also shown in the barplot at the top; the percentage of data associated with each configuration is shown in the barplot on the right. The configurations used in modelling are coarser and only consider the presence and absence of Fe and K. The map between these and the ones in the plot is as follows: $(\text{Fe},\text{K}) = \{C\}$, $(\overline{\text{Fe}},\text{K}) = \{A, F, G, J\}$, $(\text{Fe},\overline{\text{K}}) = \{D, E\}$, $(\overline{\text{Fe}},\overline{\text{K}}) = \{B, H, I\}$, where absence is denoted with a bar over the chemical element.