Chemometrics and Intelligent Laboratory Systems 150 (2016) 51–57

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Classification of multiple Dirichlet observations under a Multinomial Model

Daan de Waal [a], Roelof Coetzer [a,b], Sean van der Merwe [a,*]

[a] University of the Free State, Box 339, Bloemfontein 9300, South Africa
[b] Sasol Technology Research and Development, Private Bag 1, Sasolburg 1947, South Africa

ARTICLE INFO

Article history:
Received 20 April 2015
Received in revised form 25 October 2015
Accepted 26 October 2015
Available online 10 November 2015

Keywords:
Coal gasification
Compositional data
Dirichlet distribution
Multinomial Model

ABSTRACT

The amount of gas produced from a coal gasification facility depends crucially on the properties and the size distribution of the coal being used in the process. The particle size distribution and the composition of the coal are measured as compositional data. In this paper we apply the Dirichlet distribution for the inputs and present a new classification scheme for yielding low, medium and high gas production. The approach presented is a linear partitioning of the Dirichlet simplexes that can also be extended to high dimensional cases. An alternative clustering approach based on a distance measure is also presented.

1. Introduction

Sasol's coal-to-liquids facility delivers nearly 29% of the fuel requirements in South Africa, and the continuous improvement and optimization of the coal gasification plant are of critical importance to the company to ensure stable supply of high quality synthesis gas to the downstream units. The amount of gas produced from the coal gasification facility depends crucially on the properties and the size distribution of the coal being used in the process. Therefore, in order to optimize gas production, the relationship between the properties of the coal and the gas produced must be understood and quantified [1]. In this paper we develop and discuss a new classification scheme whereby we discriminate between coal properties and size distributions for yielding high gas production.

The particle size distribution (PSD) and the composition of the coal are measured as compositional data. The Dirichlet distribution has been widely accepted in literature for modelling compositional data, subject to the constraint that all the correlations between variables are negative [2]. The coal PSD and composition conform to the negative correlation constraint, and therefore, we fit the Dirichlet distribution to these input variables. A wider class of distributions which allows for positive correlations defined on the same sample space is the Logistic-Normal (LN) [3]. However, the LN distribution has many parameters to estimate due to the unknown covariance parameter matrix. In contrast, the Dirichlet distribution has only p unknown parameters to estimate for p compositional variables.

The current study is motivated by a coal gasification industrial problem where the amount of gas produced is a function of the quality of the coal being fed to the reactors, and it is crucially important to the business to know which coal sizes and compositions will yield optimal or sub-optimal production. The coal-to-liquids production facility gasifies no less than 40 million t of bituminous coal per year which is delivered to the factory from 5 to 7 coal sources. Therefore, the coal sizes and compositions can be very different from source to source and from week to week. Furthermore, the coal blending schedule is updated weekly subject to feed availability. Since the coal gasification process is a continuous process and the performance of the reactors is monitored in real time, it is also desired to employ a real time monitoring model for the coal feed to the reactors using data obtained from appropriate analytical equipment. Therefore, the classification scheme is presented in this study. This classification scheme can, for example, be implemented for real-time monitoring of the coal property composition and the size distribution of the coal being fed to the reactors for early detection of coal blends which may yield low gas production [4]. Real time detection of sub-optimal feed allows the factory to be pro-active in changing the coal blends to minimise the losses in gas production and to sustain profits.

In this study, three size fractions (expressed as proportions) x_1, x_2 and x_3 and three compositions (also expressed as proportions) c_1, c_2 and c_3 were considered. The dependent variable (Y) depending on covariates $C = (C_1, C_2, C_3)$ and $X = (X_1, X_2, X_3)$ is divided into three categories namely high (H), standard (S) and low (L), representing 20%, 68% and 12% of production respectively. These boundaries were chosen to illustrate the methodology only, and have no relevance to current operational performance. We assume that Y conditional on $C = \mathbf{c}$ or $X = \mathbf{x}$

* Corresponding author.
E-mail address: vandermerwes@ufs.ac.za (S. van der Merwe).

http://dx.doi.org/10.1016/j.chemolab.2015.10.012
0169-7439/

is distributed Multinomial $(n, p_{\mathbf{x}})$ where $p_{\mathbf{x}} = (p_H, p_S, p_L)$ represents the Multinomial probabilities. The aim of this paper is to classify a coal sample as H, S or L under the desired constraint $p_0 = (0.20, 0.68, 0.12)$, given its composition $(\mathbf{c})$ and sizes $(\mathbf{x})$. Here P_X denotes the random variable and $p_{\mathbf{x}}$ the observed proportion given $\mathbf{x}$ (or $\mathbf{c}$).

The classification scheme based on a single covariate (X) can be summarized as follows:

1. $Y|P_X \sim f(\mathbf{y}|P_X) \equiv \text{Multinomial}(n, P_X)$
2. $P_X|\mathbf{x} \sim k(p_{\mathbf{x}}|g(X))$ where k is some selected distribution of P_X and $g(X)$ refers to the information on X used to specify k.
3. $X|\boldsymbol{\alpha} \sim h(\mathbf{x}|\boldsymbol{\alpha}) \equiv \text{Dirichlet}(\boldsymbol{\alpha})$
4. The classification scheme: Partition the Dirichlet sample space into subspaces H, S and L such that $p_x = (p_H, p_S, p_L)$ coincides with p_0.

In this paper we demonstrate and discuss the application of the above classification scheme using simulated data from the specified probability distributions. The results are generated through simulation since the integrals cannot be solved analytically. We illustrate and discuss the performance of two alternative classification schemes:

1. a linear partitioning of the Dirichlet sample spaces, and
2. a clustering approach based on minimising a chosen metric.

The paper is outlined as follows: In Section 2, a brief introduction on applicable Dirichlet properties is given. In Section 3 the linear method is described and applied on a simulated dataset. In Section 4 the proposed method is extended to higher dimensions. In Section 5 the clustering approach is discussed. Conclusions and future work are discussed in Section 8.

2. The Dirichlet distribution

Wilks [5] and de Groot [6] provided detailed discussions on many of the properties of the Dirichlet distribution. We only mention a few. If X of order $p \times 1$ is distributed Dirichlet $(\alpha_1, \ldots, \alpha_{p+1})$, denoted by $D(\boldsymbol{\alpha})$, then the joint density is given by

$$f(\mathrm{x}) = \frac{\prod_{i=1}^{p+1}\Gamma(\alpha_i)}{\Gamma(\alpha_0)}\left\{\prod_{i=1}^{p} x_i^{a_i-1}\right\}\left\{1-\sum_{i=1}^{p} x_i\right\}^{\alpha_{p+1}-1}, \quad 0<x_i<1, \sum_{i=1}^{p} x_i<1, \quad \alpha_0 = \sum_{i=1}^{p+1}\alpha_i. \tag{1}$$

Aitchison [7] calls this distribution the Compositional Dirichlet defined on the specified simplex.

1. The means and covariances are

$$E(X_i) = \mu_i = \frac{\alpha_i}{\alpha_0}, i = 1, \ldots, p$$
$$\sigma_{ij} = \begin{cases} \dfrac{-\alpha_i\alpha_j}{\alpha_0^2(1+\alpha_0)}, i \neq j \\ \dfrac{\alpha_i(\alpha_0-\alpha_i)}{\alpha_0^2(1+\alpha_0)}, i = j. \end{cases} \tag{2}$$

   $\mu = (\mu_1, \ldots, \mu_p)$ denotes the mean of the distribution and $\Sigma = (\sigma_{ij})$ $i, j = 1, \ldots, p$ the covariance matrix.
2. The marginal distribution of X_i is Beta$(\alpha_i, \alpha_0 - \alpha_i)$ and the conditional distribution of X_i given $X_{j \neq i} = x_j, j = 1, \ldots, (i-1), (i+1), \ldots, p$ is a scaled Beta(α_i, α_{p+1}), $0 < x_i < 1 - \sum_{j \neq i}^{p} x_j$ or $\frac{X_i}{1-\sum_{j\neq i}^{p} x_j}$ is distributed Beta(α_i, α_{p+1}). The conditional distribution is useful to simulate Dirichlet observations using the Gibbs sampler. The marginal distribution of any subset J of X is again Dirichlet$(\alpha_j, \alpha_0 - \sum_{i \in J}\alpha_i)$.
3. Note in Eq. (2) that if $\boldsymbol{\alpha}$ is a multiple of $\boldsymbol{\beta}$ for two Dirichlet distributions $D(\boldsymbol{\alpha})$ and $D(\boldsymbol{\beta})$, the means are the same, but the covariance matrices differ.
4. The negative differential entropy $L = E(\log f(X))$ of the Dirichlet distribution (Eq. (1)) is given by

$$L = \log\Gamma(\alpha_0) - \sum_{i=1}^{p+1}\log\Gamma(\alpha_i) - \sum_{i=1}^{p+1}(\alpha_i - 1)\{\psi(\alpha_0) - \psi(\alpha_i)\} \tag{3}$$

   Hokela [8]. $\psi(.)$ refers to the digamma function.
5. The log Jeffreys prior [9] for the Dirichlet$(\boldsymbol{\alpha})$ is

$$0.5\sum_{l=1}^{m+1}\log\psi'(k_l) + 0.5\log\left[1 - \psi'(k_0)\sum_{l=1}^{m+1}\frac{1}{\psi'(k_l)}\right]. \tag{4}$$

   $\psi'(.)$ refers to the trigamma function.
   This prior is used later to obtain a posterior distribution of parameters. The derivation of the above is given in Van der Merwe and de Waal [10].
6. Suppose we have two Dirichlet populations, $D(\alpha_1, \ldots, \alpha_p, \alpha_{p+1})$ (parent population) and $D(\beta_1, \ldots, \beta_p, \beta_{p+1})$, then the Kullback–Leibler measure of divergence between the two follows from Eq. (3) as

$$KL = E\log\frac{f(X)}{f_0(X)} = \log\frac{\Gamma(\alpha_0)}{\Gamma(\beta_0)} - \sum_{i=1}^{p+1}\log\frac{\Gamma(\alpha_i)}{\Gamma(\beta_i)} - \sum_{i=1}^{p+1}(\alpha_i - \beta_i)\{\psi(\alpha_0) - \psi(\alpha_i)\}. \tag{5}$$

   The size of $KL \geq 0$ gives a measure of divergence between f_0 and f.

3. Classification of Dirichlet sample spaces

A dataset is constructed on compositions (C) and sizes (X) by simulating 200 Dirichlet$(\boldsymbol{\alpha})$ and Dirichlet$(\boldsymbol{\beta})$ observations independently. To each random vector of observations (C_1, C_2, X_1, X_2), the production level H, S or L is assigned according to the linear functions partitioning the two sample spaces (see Fig. 1) such that 20% of the observations are classified as H, 68% as S and 12% as L. Note that these boundaries were chosen to illustrate the methodology only, and have no relevance to current operational performance.

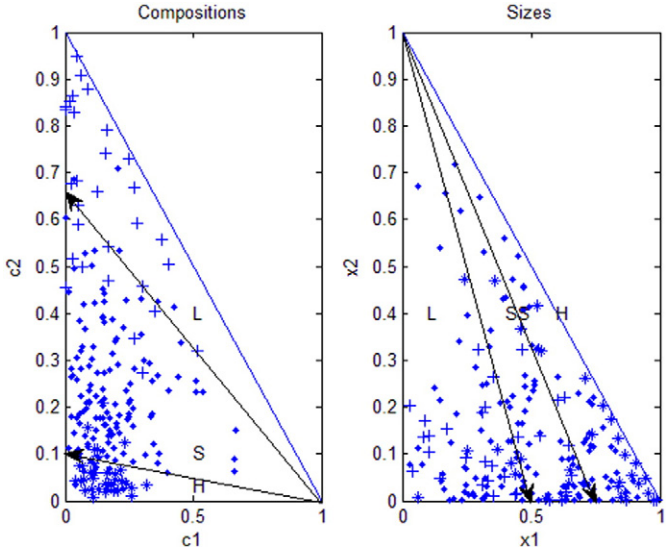

Fig. 1. Simulated Dirichlet observations on C and X with linear lines classifying each observation as H (stars), S (dots) or L (plusses).