

KAOLIN QUALITY DETERMINATION THROUGH AN ALGORITHM BASED ON NON-PARAMETRIC FUZZY LOGIC

DETERMINACIÓN DE CALIDAD DE CAOLÍN MEDIANTE UN ALGORITMO BASADO EN LÓGICA DIFUSA NO PARAMÉTRICA

CELESTINO ORDÓÑEZ

Doctor en Ingeniería de minas, Universidad de Vigo, Campus Lagoas-Marcosende, Vigo, Spain, cgalan@uvigo.es

ÁNGELES SAAVEDRA

Doctora en Matemáticas, Universidad de Vigo, Campus Lagoas-Marcosende, Vigo, Spain, saavedra@uvigo.es

MARÍA ARAÚJO

Doctor en Ingeniería de minas, Universidad de Vigo, Campus Lagoas-Marcosende, Vigo, Spain, maraujo@uvigo.es

EDUARDO GIRÁLDEZ

Doctor en Ingeniería de minas, Universidad de Vigo, Campus Lagoas-Marcosende, Vigo, Spain, edugiraldez@gmail.com

Received for review June 11th, 2010, accepted March 10th, 2011, final version March, 14th, 2011

ABSTRACT: In this article we describe a new fuzzy supervised classification method that is a modification of the fuzzy pattern-matching multidensity classifier. The latter has been demonstrated to be one of the most effective classifiers for non-convex classes. Implementing a non-parametric density estimator in one stage of the parametric method, we developed a fuzzy non-parametric classifier that manages to avoid some of the problems associated with the parametric method. The method was applied to a mineralogy problem consisting of classifying kaolin samples according to different ceramic quality levels. Our results produced error percentages that were lower than those for the parametric method.

KEYWORDS: classification, fuzzy set, non-parametric fuzzy logic, kaolin quality

RESUMEN: En este artículo se describe un método de clasificación supervisado que es una modificación del clasificador multidensidad de ajuste de patrones difuso. Se ha demostrado que este último es uno de los clasificadores más efectivos para clases no convexas. Implementando un estimador de densidad no paramétrico en una etapa del método paramétrico hemos desarrollado un clasificador no paramétrico difuso que evita algunos de los problemas asociados al método paramétrico. El método se ha aplicado a un problema de mineralurgia consistente en clasificar muestras de caolín de acuerdo con diferentes niveles de calidad cerámica. Nuestros resultados producen porcentajes de error que son inferiores a los obtenidos mediante el método paramétrico.

PALABRAS CLAVE: clasificación, conjunto difuso, lógica difusa no paramétrica, calidad de caolín

1. INTRODUCTION

In general terms, the discrimination or classification problem can be described in terms of a set of statistical techniques that enable us to study differences between populations or classes, established a priori and not necessarily exclusive. These techniques classify and assign elements to classes for which specific characteristics are known. Classification is said to be supervised when learning is implemented using a set of pre-classified elements. The aim is to predict the class to which a new element not included in the initial training set belongs. A popular procedure aimed at achieving this goal consists of adjusting the

classification functions by using a training set and then testing the percentage of correct classifications using a test set. Both the training and test sets are composed of pre-classified elements from an initial sample.

The fact that many classification methods have been developed in recent years complicates the selection of a method suitable for a given classification problem (see [1,2] for a description of traditional methods and how these are implemented). A supervised classification method that uses the fuzzy set theory during training and/or a subsequent operation is called a *fuzzy classification method* [3]. These classification methods typically apply fuzzy set theory as described by Zadeh

[4] (see [5,6] for an introduction and description of different fuzzy classification methods). It is important to point out that some of the fuzzy methods mentioned in the references given above refer to non-fuzzy methods as particular cases.

In this article we describe a non-parametric variant of the fuzzy pattern-matching multidensity classifier developed by Devillez [7] that we used to classify kaolin samples into 3 ceramic quality grades on the basis of 10 variables reflecting chemical composition. Our aim was to determine kaolin quality on the basis of values for explanatory variables obtained from a chemical analysis.

2. METHODS

2.1.1 Fuzzy supervised classification

Take a sample $\{x_1, x_2, \dots, x_n\}$ composed of n elements, for which a random p -dimensional variable (X_1, X_2, \dots, X_p) is observed. Each sample element x_i is a vector of the form:

$$x_i = (x_{i,1}, \dots, x_{i,j}, \dots, x_{i,p}) \in \mathfrak{R}^p \quad (1)$$

$$i = 1, 2, \dots, n$$

Let us assume, furthermore, that each of the elements is classified into one of the q predetermined classes C_1, C_2, \dots, C_q making up the classification space. Allocation of a new element x from the classification space to one of the q classes C_1, C_2, \dots, C_q is the main goal of a classification problem that has been handled differently in recent years. A typical procedure in many of the statistical methods used to resolve this kind of problem is to determine q membership functions $\mu_i; i = 1, 2, \dots, q$ that quantify degree of membership to different classes. A new element from the classification space, x , is allocated to the class C_i for which it obtains the greatest score; in other words: $\mu_i(x) \geq \mu_j(x); j = 1, 2, \dots, q$.

Supervised fuzzy classification methods have as their goal the calculation of the membership functions for each class. The sample is divided into two sub-samples, one for training $\{x_1, x_2, \dots, x_N\}$ and another for testing $\{x_1, x_2, \dots, x_M\}$ with the training set used in the learning phase. The learning phase consists of selecting a set of membership functions $\{\mu_1(x), \mu_2(x), \dots, \mu_q(x)\}$ that are optimal from the point of view of classification. Finally, the test sample is used to validate the accuracy of the procedure.

The classical fuzzy pattern-matching algorithm first described by Dubois et al. [8] was adapted by Devillez [7] as a fuzzy pattern matching multidensity classifier. This classifier, which has been demonstrated to be one of the most effective classifiers for classes with non-convex shapes, is implemented as described below.

2.1.1 Training

Each of the classes $C_k, k = 1, 2, \dots, q$ that make up the classification space are divided into a number s_k of sub-classes using the fuzzy c -means algorithm [9,10] which minimizes the within-class sum of squared errors in the following conditions:

$$\sum_{l=1}^{s_k} m_{i,l} = 1, \sum_{i=1}^N m_{i,l} > 0, m_{i,l} \in [0,1] \quad (2)$$

$$i = 1, 2, \dots, N, l = 1, 2, \dots, s_k$$

where $m_{i,l}$ are the membership functions to be determined. The objective function to be minimized is given by the following expression:

$$J = \sum_{i=1}^N \sum_{l=1}^{s_k} m_{i,l}^r d^2(x_i, c_l) \quad (3)$$

where N is the number of data items in the training subset, c_l is the vector representing the centroid of the l -th sub-class, $C_{k,l}$ of the class C_k , r is the fuzzy exponent, which should, strictly speaking, be greater than 1, and where $d(\bullet, \bullet)$ is a measure of the distance between two points calculated as:

$$d(x_i, c_l) = \sqrt{(x_i - c_l)' A (x_i - c_l)} \quad (4)$$

where A is the distance norm matrix, identity matrix for the Euclidean distance or the inverse of the variance-covariance matrix for the Mahalanobis distance.

Minimizing the objective function J offers the membership functions for each sub-class as a solution:

$$m_{i,l} = \frac{d^{2(r-1)}(x_i, c_l)}{\sum_{j=1}^{s_k} d^{2(r-1)}(x_i, c_j)} \quad (5)$$

$$i = 1, 2, \dots, N, l = 1, 2, \dots, s_k$$

This procedure can be implemented, for example, using the FuzME program [11].

Following the application of the algorithm, a set $\{C_{k,l}\}_{l=1}^{s_k}$ of sub-classes is obtained for each class C_k . The elements in the training set $\{x_1, x_2, \dots, x_N\}$ are allocated to the sub-class with the maximum membership function as calculated by the fuzzy algorithm.

Learning continues with the calculation of a membership function for each sub-class and each variable X_j . These functions can be calculated on the basis of the probability estimated from the histograms for the data. In other words, given a sub-class $C_{k,l}$ and a variable X_j , the histogram is calculated using the values $x_{i,j}$ of the elements x_i from the training set that have been allocated to the sub-class $C_{k,l}$.

Determined during the training process are $p \times \left(\sum_{k=1}^q s_k \right)$ histograms that have to be transformed into fuzzy membership functions. The membership functions can be deduced from the probability distributions associated with the calculation of the histograms using any of the transformations proposed by Dubois et al. [12]. Following Devillez [7], we implemented one of the first probability-possibility transformations, introduced by Dubois and Prade [13], who define the probability distribution associated with a histogram as the discrete distribution given by $\{y_1, y_2, \dots, y_h\}$ —which are central values in the histogram bins—and by the probabilities $\{p_1, p_2, \dots, p_h\}$ —calculated as the ratio between bin height and the sum of all the heights. The probability values are next arranged in descending order: $p_{(1)} \geq p_{(2)} \geq \dots \geq p_{(h)}$. The values for the membership function associated with the values $\{y_1, y_2, \dots, y_h\}$ are then calculated:

$$\pi_{(i)} = i \times p_{(i)} \sum_{j=i+1}^h p_{(j)}, i = 1, 2, \dots, h \tag{6}$$

The membership function is completed by means of a linear interpolation of the above values.

The learning phase concludes with $p \times \left(\sum_{k=1}^q s_k \right)$ membership functions, which we can denote as:

$$\mu_{j,k,r}^1; j = 1, 2, \dots, p, \tag{7}$$

$$k = 1, 2, \dots, q, r = 1, 2, \dots, s_k$$

2.1.2 The nonlinear classification rule

For an element $x_i = (x_{i,1}, \dots, x_{i,j}, \dots, x_{i,p}) \in \mathfrak{R}^p$ from the test sample $\{x_1, x_2, \dots, x_M\}$ the degree of membership to each sub-class $C_{k,l}$ is evaluated by applying the minimum operator to the membership functions calculated in the training phase:

$$\eta_{k,l}^1(x_i) = \min \left\{ \mu_{1,k,l}^1(x_{i,1}), \dots, \mu_{p,k,l}^1(x_{i,p}) \right\} \tag{8}$$

$$k = 1, 2, \dots, q, l = 1, 2, \dots, s_k$$

Next, the degrees of membership to each sub-class are combined, resulting in a degree of membership to each class $C_k; k = 1, 2, \dots, q$ as follows:

$$\mu_k^1(x_i) = \min \left\{ 1, \sum_{l=1}^{s_k} \eta_{k,l}^1(x_i) \right\} \tag{9}$$

$$k = 1, 2, \dots, q$$

The element x_i is allocated to the class for which its degree of membership is highest.

2.2 The non-parametric fuzzy supervised classification method

The fuzzy pattern-recognition multidensity method described above estimates, in the learning stage, probability functions from the histograms calculated with the values $x_{i,j}$ of the training set elements x_i allocated to the sub-classes $C_{k,l}$. Although the use of the histogram is very convenient for graphically representing the empirical distribution of frequencies, it has some drawbacks as a method for estimating the theoretical distribution of frequencies. For continuous variables, kernel density estimation methods are often preferable.

The kernel density estimators belong to a class of estimators called *non-parametric density estimators*. Unlike the parametric estimators, which require a fixed functional form, the non-parametric estimators do not have a fixed structure but depend exclusively on the set of observations for the density being estimated. (See Härdle et al. [14] for a detailed explanation of these methods.)

The rationale behind the use of the non-parametric statistical techniques can be found in the fact that the width and the initial points of each histogram bar needs to be determined when calculating a histogram;

however, this leads to a problem in that the histograms depend on parameters that are often selected arbitrarily. These drawbacks inspired the development of the first kernel estimators [14].

Therefore, applying a kernel density estimator, the elements in the training sample can be used in order to non-parametrically estimate the density function of the variable X_j for a sub-class $C_{k,l}$:

$$\hat{f}(y) = \frac{1}{N_{k,l}h} \sum_{i=1}^{N_{k,l}} K\left(\frac{y - x_{i,j}}{h}\right) \quad (10)$$

where $N_{k,l}$ represents the number of elements in the training sample allocated to sub-class $C_{k,l}$, h is the smoothing parameter, and K is the kernel function. This function should confirm that $\int K(t)dt = 1$ in order to ensure that the density function \hat{f} integrates to 1 and where the kernel function K is usually chosen to be a smooth unimodal function with a peak at 0.

Note that, in order to eliminate dependence on the initial points of the bins, kernel estimators centre a kernel function at each data point and smooth out the contribution of each observed data point over a local neighbourhood for that data point. The extent of this contribution is dependent upon the shape of the kernel function adopted and the smoothing parameter.

Implementation of this variant in the training process gives rise to $p \times \left(\sum_{k=1}^q s_k\right)$ density functions from which the fuzzy membership functions are deduced, using either of two alternatives: discretize the estimated density function \hat{f} and apply a probability-possibility transformation similar to that referred to above, or implement a continuous probability-possibility transformation using one of the procedures described in Dubois et al. [11]. Obtained in both cases are $p \times \left(\sum_{k=1}^q s_k\right)$ membership functions, which we can now denote as:

$$\mu_{j,k,l}^2; j = 1, 2, \dots, p; k = 1, 2, \dots, q; \\ l = 1, 2, \dots, s_k$$

In the nonlinear classification rule, the membership functions $\mu_{j,k,l}^2$ are evaluated in the coordinates of the elements of the test set, resulting in the classification

functions $\eta_{k,l}^2$ and v_k^2 , and in a procedure that is identical to that described in the previous section.

$$\eta_{k,l}^2(x_i) = \min\left\{\mu_{1,k,l}^2(x_{i,1}), \dots, \mu_{p,k,l}^2(x_{i,p})\right\} \\ k = 1, 2, \dots, q; l = 1, 2, \dots, s_k$$

$$\mu_k^2(x_i) = \min\left\{1, \sum_{l=1}^{s_k} \eta_{k,l}^2(x_i)\right\} \quad (11)$$

$$k = 1, 2, \dots, q$$

2.3 The data

With a view to conducting a comparative study, the two classification methods were applied to a set of real data. The data used for the application of these techniques were obtained from analyses of samples taken from a primary kaolin deposit in one of the most important kaolin mining areas of Galicia (NW Spain), namely, Vimianzo (A Coruña). The samples were wet-sieved and mineralogical and chemical analyses were performed for 50 μ -grained fractions following standard procedures. The following information was obtained from the analyses: mineralogical data (percentages of quartz, mica, feldspar, and kaolinite), chemical data (SiO₂/Al₂O₃ ratio, percentages of Fe₂O₃, TiO₂, K₂O, MgO, and loss on ignition [LOI]). These parameters are those typically used to identify the suitability of a sample for use in ceramics manufacturing.

All the parameters ($p = 10$) were selected as variables to which the techniques apply. For each variable, three quality classes were defined—top quality, medium quality, and poor quality [16]—, according to quality rankings as determined in the literature for the ceramic industry [17,18]. In addition to these numerical variables, a ceramic quality variable was included, also defined in terms of three classes reflecting the saleability of the sample—top quality, medium quality and poor quality.

Ceramic quality was evaluated for 1410 samples on the basis of these mineralogical, chemical, and physical parameters. The sample was divided into a training set ($n = 558$) and a test set. The number of classes considered for the classification of the sample elements was $q = 3$, in accordance with the ceramic quality grades (top, medium, and poor) defined above.

2.4 RESULTS

Figure 1 shows the histogram of the kaolin percentage values corresponding to one of the sub-classes in which the medium quality class was divided. Superimposed on the histogram is the membership function obtained by the supervised fuzzy classification method.

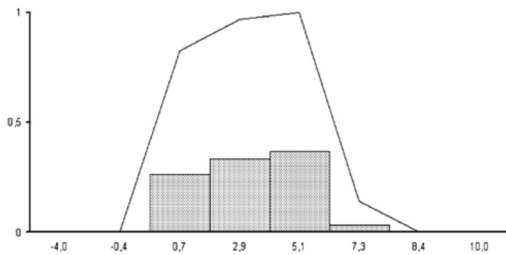


Figure 1. Histogram and membership function were the supervised fuzzy classification method applied to the kaolin percentage variable (x-axis) and one of the medium quality sub-classes for ceramics

The fuzzy classification method implemented in this research used the density of a Gaussian distribution as the function K . The smoothing parameter h was chosen using the smoothed bootstrap method proposed by Cao [19]. This method for selecting the parameter is particularly suitable when working with independent samples. It was found that greater percentages of correct classifications were obtained using half of the smoothing parameter.

Figure 2 shows the density function when estimated non-parametrically and, for the same variable and sub-class, the membership function obtained by means of the non-parametric fuzzy supervised classification supervised methods.

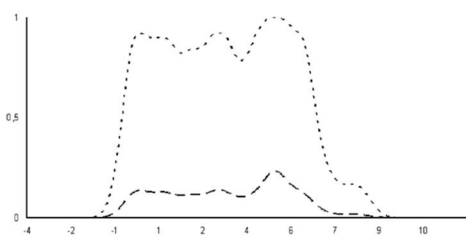


Figure 2. The non-parametric density estimation (broken line) and the membership function obtained by the non-parametric fuzzy supervised classification method (dotted line) for the kaolin percentage variable (x-axis) and one of the medium quality sub-classes

The percentage of correctly classified elements in the training set was 99.64 % for the fuzzy supervised method and 99.82 % for the non-parametric method, whereas the same percentages of the validation test set were 98.01 % and 99.10% , respectively.

3. CONCLUSIONS

This article describes a non-parametric fuzzy classification method that is a variant on the fuzzy pattern matching multidensity method. The difference between the two approaches lies in the method for determining the density function; our method uses the histogram of the data, whereas the multidensity method uses a kernel function. The main advantage of the non-parametric method is that it does away with some of the drawbacks associated with the construction of the histogram, such as the problem of defining the width and initial point for each interval in the histogram.

Our method was tested in determining the quality of kaolin samples obtained from a kaolin deposit for ceramic manufacturing purposes. The level of correct classification was quite similar for both methods, although there was a slight improvement in the fuzzy classification methods.

As future research, we will explore the possibility of automatically selecting the smoothing parameter so that the non-parametric fuzzy supervised classification method becomes genuinely competitive with that proposed by Devillez [7].

ACKNOWLEDGMENTS

This research has been partially supported by grant PGIDIT06PXIC300117PN of the *Xunta de Galicia* and also by grant PGIDIT07PXIB300191PR of the *Xunta de Galicia* for the first author.

REFERENCES

- [1] Härdle, W., *Multivariate Statistical Analysis*, Springer, Berlin, 2003.
- [2] Izenman, A.J., *Modern multivariate statistical techniques regression, classification and manifold learning*, Springer, New York, 2008.

- [3] Jiménez, C., Soto, C. Aprendizaje supervisado para la discriminación y clasificación difusa, *Dyna* 169, pp. 26-33, 2011.
- [4] Zadeh, L.A., Fuzzy sets, *Information and Control*, 8, pp. 338-353, 1965.
- [5] Kuncheva, L.I., Fuzzy classifier design, *Physica Verlag*, 2000.
- [6] Dumitrescu, D., Lazzerini, B. and Jain, L.C., Fuzzy sets and their application to clustering and training, *CRC Press*, Boca Raton, 2000.
- [7] Devillez, A., Four fuzzy supervised classification methods for discriminating classes of non-convex shape, *Fuzzy sets and systems*, 141(2), pp. 219-240, 2004.
- [8] Dubois, D., Prade, H. and Testemale, C., Weighted fuzzy pattern matching, *Fuzzy sets and systems*, 28(3), pp. 313-331, 1988.
- [9] Pedrycz, W. and Vulkovich, G., Fuzzy clustering with supervision, *Pattern Recognition*, 37(7), pp. 1339-1349, 2004.
- [10] Bezdek, J.C., Erlich, R. and Full, W., FCM: The fuzzy c-means clustering algorithm, *Computers and Geosciences* 10(2), pp. 191-203, 1984.
- [11] Minasny, B. and Mc Bratney, A.B., FuzME version 3.0, Australian Centre for Precision Agriculture, The University of Sydney, Australia, 2002.
- [12] Dubois, D., Fulloy, G., Mauris, H. and Prade, H., Probability-possibility transformations, triangular fuzzy sets and probabilistic inequalities, *Reliable Computing* 10, pp. 273-297, 2004.
- [13] Dubois, D. and Prade, H., *Théorie des possibilités, Application à la représentation des connaissances en informatique*, Masson, Paris, 1987.
- [14] Härdle, W., Müller, M., Sperlich, S. and Werwatz, A., *Nonparametric and semiparametric models*, Springer, New York, 2004.
- [15] Parzen, E., On estimation of a probability density function and mode, *Annals of Mathematical Statistics* 33(3), pp. 1065-1076, 1962.
- [16] Taboada, J., Rivas, T., Araujo, M. and Argüelles, A., A fuzzy expert system application to the evaluation of ceramic- and paper-quality kaolin. *Applied Clay Science*, 33, pp. 287-297, 2006.
- [17] Prasad, M.S., Reid, K.J. and Murria, H.H. Kaolin: processing, properties and applications. *Applied Clay Science*, 6-3,4, pp. 87-119, 1991.
- [18] Konta, J., Clay and man: clay raw materials in the service of man, *Applied Clay Science*, 10(4), pp. 275-353, 1995.
- [19] Cao, R., Bootstrapping the Mean Integrated Squared Error, *Journal of Multivariate Analysis* 45(1), pp. 137-160, 1993.