Continuous Multi-way Shape Measure for Dissimilarity Representation

Diana Porro-Muñoz[1,2], Robert P.W. Duin[2],
Mauricio Orozco-Alzate[3], and Isneri Talavera Bustamante[1]

[1] Advanced Technologies Application Center (CENATAV), Cuba
{dporro,italavera}@cenatav.co.cu
[2] Pattern Recognition Lab, TU Delft, The Netherlands
r.duin@ieee.org
[3] Universidad Nacional de Colombia - Sede Manizales, Colombia
morozcoa@bt.unal.edu.co

Abstract. For many applications, a straightforward representation of objects is by multi-dimensional arrays e.g. signals. However, there are only a few classification tools which make a proper use of this complex structure to obtain a better discrimination between classes. Moreover, they do not take into account context information that can also be very beneficial in the classification process. Such is the case of multi-dimensional continuous data, where there is a connectivity between the points in all directions, a particular (differentiating) shape in the surface of each class of objects. The dissimilarity representation has been recently proposed as a tool for the classification of multi-way data, such that the multi-dimensional structure of objects can be considered in their dissimilarities. In this paper, we introduce a dissimilarity measure for continuous multi-way data and a new kernel for gradient computation. It allows taking the connectivity between the measurement points into account, using the information on how the shape of the surface varies in all directions. Experiments show the suitability of this measure for classifying continuous multi-way data.

Keywords: Classification, Continuous multi-way data, Dissimilarity representation, Object representation.

1 Introduction

Representation of objects by matrices or higher-order arrays has become very popular in many application areas. These multi-dimensional structures of objects can be created for some specific purpose or they can be obtained directly from acquisition equipments e.g. excitation-emission autofluorescence measurements and time-frequency representation of signals.

Tools that make a proper use of these multi-way structures are needed. Traditional multivariate methods are not suitable for it. Data would have to be re-arranged in a vector, thus the information of the multi-way structure is lost

and huge-dimensional problems are created. Several methods have been proposed for multi-way data analysis, mainly in the psychometrics and chemometrics fields. However, most of them aim for regression or exploratory analysis [1]. The development of tools for multi-way classification is rather poor in comparison with the large amount of methods for other purposes. There are basically three main multi-way classification approaches, namely: NPLS-DA, traditional classifiers on the scores of multi-way decomposition methods like PARAFAC [2], and NSIMCA [3]. These methods succeed in making use of the multi-way structure. However, this might not be enough for a proper analysis of the data. Continuous data, for example, are just numerically analyzed as a set of sampled, individual features. Features are in this paper the measurement directions of a multi-way data set. Important discriminative information, like connectivity between features and their shape is not considered in the analysis.

Recently, the Dissimilarity Representation (DR) approach was introduced for multi-way data classification [4]. The DR approach consists in representing objects by their proximities with respect to other objects [5]. As classes of objects are determined by how (dis)similar they are, the authors advocate that for classification, a representation based on dissimilarities between objects may be more informative than the traditional feature based representation.

One of the goals and advantages of this approach is the possibility of introducing discriminative context information in the representation of objects by the dissimilarity measure. In the case of 2D or any-dimensional continuous data, e.g. spectroscopic data, it could be important for their analysis to consider their continuous nature. Several 2D measures have been proposed for image comparison; however most of them are just based on pairwise comparison of objects, ignoring the continuous nature of images [6]. Recently, the 2Dshape [4] measure was introduced for this purpose. In order to reflect the continuous shape of the data, comparison of objects is based on the differences between the first Gaussian derivatives in each direction. However, it is based on the combination of 1D dissimilarities; hence it does not analyze the combined 2D shape changes [1].

In this paper, we propose a new dissimilarity measure, the Continuous Multi-way Shape (CMS) measure, that exploits the information on the whole structure of multi-dimensional continuous data. It is based on differences between gradients of objects, so shape changes of the surfaces are considered. The computation of the gradient components is usually based on convolution kernel operators. We introduce a gradient kernel operator where each partial derivative is computed as the derivative of a polynomial fitted to the analyzed points. The proposed measure will be compared to 2Dshape [4], and to two 2D measures [6], which are not designed for continuous data. It is shown that considering the continuous nature of data can be beneficial to improve its classification.

The paper is organized as follows. Fundamentals of the dissimilarity representation approach are presented in the first part of Section 2. In the second part, the new measure and its generalization are detailed. Experiments and results are discussed in Section 3. Conclusions are presented in Section 4.

---

[1] The naming of the procedure given in paper [4], 2Dshape, can be confusing.

2   Dissimilarity Representation for Multi-way Data

In the DR [5] approach, new features are defined for the objects, such that they are represented by their proximities to a set of representative objects of each class. The fact (property) that dissimilarities should be smaller for similar objects and larger for different ones, suggests that they could be used as more discriminative features, if a suitable measure is used.

Thus, in this approach, given a set of training objects $\mathbf{X} = \{x_1, x_2, ..., x_l\}$, a representation set (a set of prototypes for each class) $\mathbf{R} = \{r_1, r_2, ..., r_p\}$, and a dissimilarity measure; the distance between each object $x_i \in \mathbf{X}$ to each object $r_h \in \mathbf{R}$ will be defined as $d(x_i, r_h)$. The representation set $\mathbf{R}$ can be a subset of $\mathbf{X}$, $\mathbf{R} \subseteq \mathbf{X}$ or $\mathbf{X}$ itself, being then $\mathbf{D}(\mathbf{X}, \mathbf{X})$ a square dissimilarity matrix, or $\mathbf{R}$ and $\mathbf{X}$ can be completely different sets. There are a number of approaches to select prototypes of the representation set [5].

Let us assume we have a multi-way or n-dimensional array $\underline{Y} \in \mathbb{R}^{I_1 \times I_2 \times ... \times I_n}$ (training set), where object $\underline{Y_i} \in \mathbb{R}^{I_1 \times I_2 \times ... \times I_{n-1}}$. To build the dissimilarity space, the mapping $\phi(\cdot, \mathbf{R}) : \mathbb{R}^{I_1 \times I_2 \times ... \times I_{n-1}} \to \mathbb{R}^h$ is defined, such that for every object $\phi(\underline{Y_i}, \mathbf{R}) = [d(\underline{Y_i}, \underline{r_1}), d(\underline{Y_i}, \underline{r_2}), ..., d(\underline{Y_i}, \underline{r_h})]$. We need then a $(n\text{-}1)$-dimensional dissimilarity measure for its computation. Classifiers are built in this space, as in any feature space. Consequently, the relationship between all objects in the training and representation sets is used for classification. If a suitable measure is chosen, the compactness property (objects from the same class should be similar and objects from different classes should be different) of the classes should be more pronounced. Therefore, it should be easier for the classifiers to discriminate between them, since linear classifiers in the dissimilarity space may correspond to non-linear classifiers in the feature space. In general, any arbitrary classifier operating on features can be used [5].

3   Continuous Multi-way Shape Measure

A measure that somehow respects the multi-way structure of the data (considers the relationship of different directions) is needed. It should also make use of the continuity and shape information of the multi-dimensional continuous data.

Given a $n - 1$ dimensional object, each point in the multi-dimensional surface could be analyzed with a $n - 1$ dimensional window, such that shape changes in all directions can be taken into account. Thus, the comparison between two objects should be based on the differences of their multi-way shape, considering the connectivity that exists between the neighboring points in the different directions. In the case of 1D continuous data e.g. spectral data, derivatives are the commonly used tool to evaluate shape changes. For multi-dimensional functions, the gradient is the natural extension of the derivative concept.

Although data may have a continuous nature, they are captured by the sampling procedures of sensors as a collection of discrete values. As derivatives are undefined for discrete functions, they need to be estimated somehow to be used on these data. A widely used method for approximating the derivative of a discrete function is the application of linear filters by convolution.

Given the multi-dimensional discrete functions $\underline{Y}$ and $\underline{H}$, a convolution operation is defined as $\underline{Y}' = \underline{Y} * \underline{H}$, where $*$ is the convolution operator [7]. Thus, given two objects $\underline{Y_a}$ and $\underline{Y_b}$, they can be compared by computing the difference between the gradients of the surfaces that represent these objects. As derivatives and therefore the gradient are very noise sensitive, data should be smoothed before performing these operations. A common way to smooth data is by convolving it with a Gaussian filter. Following the previous ideas, the Continuous Multi-way Shape (CMS) measure is defined as:

Definition 1. *Let $\underline{Y}$ be a n-way data set and let $\underline{Y_a}$, $\underline{Y_b}$ be two objects from this data set. The dissimilarity between $\underline{Y_a}$ and $\underline{Y_b}$ can be computed as:*

$$d_G(\underline{Y_a}, \underline{Y_b}) = \left\| \sum_{i=1}^{f} \underline{Y_a} * G_\sigma * H_i - \underline{Y_b} * \underline{G_\sigma} * \underline{H_i} \right\|_F \tag{1}$$

where $\| \cdot \|_F$ is the Frobenius norm for tensors [8], $\underline{G_\sigma}$ a Gaussian convolution kernel to smooth the data first, $\underline{H_i}$ is a partial derivative kernel and f is the amount of partial derivatives in the different directions in order to obtain the gradient.

3.1 Gradient Kernels

The Prewitt and Sobel operators [7] are 2D convolution kernels based on linear filters that compute the average gradient components of three adjacent lines and columns to overcome the noise sensitivity. They differ in that central weights for the smoothing are higher. Horizontal and vertical Prewitt and Sobel kernels are then defined as follows:

$$H_x^P = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \quad H_y^P = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \quad H_x^S = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad H_y^S = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \tag{2}$$

The previous gradient operators are based on approximating a partial derivative at a point p by computing the slope of the line that fits the previous and next point of p in the direction of the derivative. Extensions of these filters to diagonal directions and to higher dimensions can be found [9], and the same idea of a line fitting is kept.

3.2 Gradient Polynomial-Based Kernel for the CMS Measure

We propose to approximate each partial derivative in point p as the derivative of the polynomial of degree t, which is obtained by interpolating p and its t nearest points in the direction of the derivative.

For a 2D kernel of size $[3 \times 3]$ this approach is equivalent to applying the Prewitt operator. So, the particular case of a 2D kernel of size $[5 \times 5]$ will be explained here. For 2D objects, we want to analyze the derivatives in four directions (horizontal, vertical and the two diagonals). Without loss of generality, let us see

the case for the horizontal direction. Assume there are five points $p_0 = (-2, y_0)$, $p_1 = (-1, y_1)$, $p_2 = (0, y_2)$, $p_3 = (1, y_3)$ and $p_4 = (2, y_4)$, so a fourth degree polynomial should be approximated to compute the derivative in each direction. We chose these values for x just for simplicity, but it really does not matter, the shape of the polynomial is the same: $P(x) = ax^4 + bx^3 + cx^2 + dx + e$. Its derivative is the cubic polynomial $P'(x) = 4ax^3 + 3bx_2 + 2cx + d$ and, evaluated in point p_2, it is reduced to $P'(0) = d$. If we solve the system of equations from evaluating all points in $P(x)$, we obtain $P'(0) = d = \frac{y_0 - 8y_1 + 8y_3 - y_4}{12}$, which can be wrapped in a filter kernel.

As for Prewitt and Sobel operators where adjacent lines are averaged, an approximated average polynomial of the adjacent set of points can be obtained. Then, the following convolution kernels for horizontal, vertical, main diagonal and secondary diagonal are defined respectively:

$$H_x^L = \begin{array}{|c|c|c|c|c|} \hline 1 & -8 & 0 & 8 & 1 \\ \hline 1 & -8 & 0 & 8 & 1 \\ \hline 1 & -8 & 0 & 8 & 1 \\ \hline 1 & -8 & 0 & 8 & 1 \\ \hline 1 & -8 & 0 & 8 & 1 \\ \hline \end{array} \quad H_y^L = \begin{array}{|c|c|c|c|c|} \hline 1 & 1 & 1 & 1 & 1 \\ \hline -8 & -8 & -8 & -8 & -8 \\ \hline 0 & 0 & 0 & 0 & 0 \\ \hline 8 & 8 & 8 & 8 & 8 \\ \hline -1 & -1 & -1 & -1 & -1 \\ \hline \end{array} \quad H_{md}^L = \begin{array}{|c|c|c|c|c|} \hline 0 & 0 & 0 & 1 & 1 \\ \hline 0 & 0 & -8 & -8 & 1 \\ \hline 0 & 8 & 0 & -8 & 0 \\ \hline -1 & 8 & 8 & 0 & 0 \\ \hline -1 & -1 & 0 & 0 & 0 \\ \hline \end{array} \quad H_{sd}^L = \begin{array}{|c|c|c|c|c|} \hline 1 & 1 & 0 & 0 & 0 \\ \hline 1 & -8 & -8 & 0 & 0 \\ \hline 0 & -8 & 0 & 8 & 0 \\ \hline 0 & 0 & 8 & 8 & -1 \\ \hline 0 & 0 & 0 & -1 & -1 \\ \hline \end{array} \quad (3)$$

The same idea can be generalized to larger windows, but higher order polynomials are used. These filters can also be extended to n-way arrays. Polynomials will be determined in the same manner according to the size of the window, but now there will be more directions to be analyzed. For example, in the case of a 4-way array where objects are 3D, if we use a $[3 \times 3 \times 3]$ window, 13 directions can be analyzed as in the following figure:
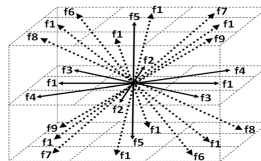

Fig. 1. The 13 directions to be analyzed in a 3D object

The proposed CMS measure can be seen as a generalization of the idea of the 2Dshape [4] measure for 2D objects. Both measures are based on smoothing surfaces with a Gaussian kernel and their partial derivatives are compared in order to take the shape of the functions into account. However, in the case of the 2Dshape measure, a 1D Gaussian derivative is computed independently for every feature in the horizontal and vertical directions of a 2D object, treating the feature as a 1D signal, without taking into account the relationship between the features in the two directions. Differences in each direction are then combined.

In contrast, the CMS measure already considers 2D information in the smoothing process by applying a 2D Gaussian kernel. Moreover, with 2D gradient kernel operators, more global information (relationship between features) is considered. As partial derivatives are computed on the average line obtained from a number of rows/columns (depending on the kernel size), information from the neighboring signals is used. The most remarkable feature of the CMS measure is that it

is not restricted to measure the shape in two directions only. Although CMS is based on the idea of the gradient of an image, which is mathematically defined by a vector of 2 components i.e. horizontal and vertical, this measure allows analyzing other directions e.g. diagonals. This way, more accurate approximations of the information on the shape of objects and dependencies in the different directions are computed for the comparison. With the introduced polynomial-based kernels, the partial derivatives can be approximated as the derivative of higher-degree polynomials instead of a simple line. Consequently, it is to be expected that the multi-way shape information can be better modeled with the proposed CMS measure, leading to a better discrimination of these types of objects. At last, the CMS measure has been defined for continuous multi-way objects in general, while 2Dshape can only be applied to 2D objects.

4 Experimental Setup and Discussion

In this section, we present experiments with a Regularized Linear Discriminant classifier based on dissimilarities. Experiments are conducted on five 2D continuous data sets of different sources. Our aim is to compare the performance of classifiers on the two shape-based measures, 2Dshape and the proposed CMS measure for multi-way continuous data. These performances will also be compared with the non-shape based measures Frobenius and Yang, which are versions of the AMD [6] distance with weights $p = 2$ and $p = 1$ respectively.

For the different data sets, experiments were carried out differently. For small data sets (Parma ham and St John's), classification errors were obtained in a 10 times k-fold cross-validation (CV). The Enzyme data has a training and test set, so the classifier is evaluated on the test set. In the case of Colon and Volcano data, 10 different training and test sets were randomly chosen and the error values were averaged. For the three bigger data sets, i.e. Enzyme, Colon and Volcano, a part of the data was used to optimize the measures parameters in a CV procedure. The rest was then used as explained before. For the other two, as they are too small, the parameters were optimized with the whole data sets.

The first data set is private and it comes from 1200 patches of 1024×1024 pixels of 36 colon tissue slides from Atrium hospital in Heerlen, The Netherlands. Patches were filtered with Laplace filters in 90 different scales using $\sigma = 2.\hat{}[0.1 : 0.1 : 9]$. The log-squares of the results are summarized in 60 bin normalized histograms with bin centres $[-50 : 1 : 9]$. Thus, a 90×60 array is obtained for every patch, leading to a three-way array of $1200 \times 90 \times 60$. The patches are labeled in two classes: Normal and Tumor. A representation set of 550 prototypes was randomly chosen from the training set. The second data set corresponds to seismic signals from Nevado del Ruiz volcano in the Colombian Andes. The data set is composed of 12032-point signals of two classes of volcanic events: Long-Period earthquakes, and Volcano-Tectonic earthquakes. A 2D time-frequency representation was computed for each signal with a 256-points (window size) Short-Time Fourier Transform (STFT), with 50% overlap. The concatenation of the obtained spectrograms results in a $470 \times 93 \times 129$ three-way array. The dissimilarity matrix has a size of 470×100.

The third and fourth data sets are from public domains and they are both obtained by Fluorescence spectroscopy. The first of them consists of a training set of size $323 \times 15 \times 15$ and a test set of $53 \times 15 \times 15$. The two variable directions correspond to excitation and emission wavelengths respectively. The classification problem consists in determining the quality (Low or High) of a process according to the enzyme activity [10]. A representation set of 100 prototypes was randomly chosen from the training set, thus the dissimilarity matrices have a size of 323×100 and 53×100 for training and test sets respectively. The other data set has a size of $67 \times 11 \times 13$ and the purpose is to determine the age range of a Parma ham sample: raw (0 months), salted (3 months), matured (11 and 12 months) and aged (15 and 18 months) [3]. The last data set consists of 108 samples of carrot juice, which have been crystallized, with the aim of describing their quality (Good/Bad) [11]. Images of size 528×528 of each biocrystallized sample were taken. Gabor filters with a bank of 128 filters from 16 orientations was applied, resulting in a four-way data set of $108 \times 528 \times 528 \times 128$. Thus, 128 dissimilarity matrices were computed with 2D measures on the 528×528 matrices of each filter and latter averaged. All objects were used as representation set in the DR.

Table 1. Classification error with different measures: CMS measure with Prewitt, Prewitt in 4 directions (including diagonals), Sobel, Sobel in 4 directions and Polynomial filter, 2Dshape measure, Frobenius and Yang

| Data | CMS | | | | | 2Dshape | No shape | |
	Prew.	Prew.(4d)	Sob.	Sob.(4d)	Polyn.		Frob	Yang
Colon cancer	11.0	11.5	11.2	12.0	**9.5**	12.7	13.3	13.3
Volcano	28.0	25.6	28.2	23.4	23.4	**20.9**	40.0	28.7
Enzyme	9.4	**5.7**	9.4	9.4	9.4	13.2	9.4	9.4
Parma ham	3.7	**2.4**	3.7	2.5	3.7	2.9	4.5	4.3
Carrot juice	**7.2**	**6.0**	**7.2**	6.3	7.1	**8.3**	9.8	10.7

Results are shown in Table 1. It can be seen that as expected, measures which take the continuous information of data into account give the best results. The CMS measure, with most filters, outperforms the results obtained with the 2Dshape measure in general, corroborating our previous analysis. The selection of the kernel to be applied should depend on the problem at hand and how rough shape changes are. Larger kernels should be able to capture better the changes in the surface when these are not so sudden. However, if there are shape changes in small regions, they might be averaged in a large window. Thus, small windows should work better in these cases. It is shown that results are improved by using the diagonal directions in the Prewitt and Sobel operators. This supports the previously discussed argument that if more directions are analyzed, there can be more information which contributes to a better discrimination of the classes.

5   Conclusions

We introduced a multi-dimensional dissimilarity measure for multi-way continuous data based on the computation of the gradient. This was proposed with the

aim of applying the DR approach as a classification tool for this type of data. The new measure allows taking into account the complex multi-dimensional structure, such that the connectivity and shape information of the surfaces (objects) can be considered in the dissimilarity representation of the objects. The way the measure has been defined, allows to use different gradient convolution kernels, according to the problem at hand. This measure was compared to the 2Dshape measure and other non-shape based measures for the classification of 2D objects. Results have corroborated the presented argument that considering the continuous multi-way nature of these types of data in their analysis can lead to better results. Moreover, it is shown that by taking into account the information in more directions, results can be improved.

References

[1] Smilde, A.K., Bro, R., Geladi, P.: Multi-way Analysis. In: Applications in the Chemical Sciences. Wiley Publisher, England (2004)
[2] Acar, E., Aykut-Bingol, C., Bingol, H., Bro, R., Yener, B.: Multiway analysis of epilepsy tensors. Bioinformatics 23, i10–i18 (2007)
[3] Durante, C., Bro, R., Cocchi, M.: A classification tool for N-way array based on SIMCA methodology. Chem. and Intell. Lab. Syst. 106, 73–85 (2011)
[4] Porro-Muñoz, D., Duin, R.P.W., Talavera, I., Orozco-Alzate, M.: Classification of three-way data by the dissimilarity representation. Signal Processing 91(11), 2520–2529 (2011)
[5] Pekalska, E., Duin, R.P.W.: The Dissimilarity Representation For Pattern Recognition. Foundations and Applications. World Scientific, Singapore (2005)
[6] Zuo, W., Zhang, D., Wang, K.: An assembled matrix distance metric for 2DPCA-based image recognition. Pattern Recognition Letters 27, 210–216 (2006)
[7] Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 3rd edn. Prentice-Hall, Inc., Upper Saddle River (2006)
[8] Lathauwer, L., De Moor, B.: From matrix to tensor: Multilinear algebra and signal processing. In: Proc. 4th Int'l Conf. on Mathematics in Signal Processing, vol. 1, pp. 1–11 (1996)
[9] Jetto, L., Orlando, G., Sanfilippo, A.: The edge point detection problem in image sequences: Definition and comparative evaluation of some 3D edge detecting schemes. In: Proc. of the 7th Mediterranean Conference on Control and Automation (MED 1999), pp. 2161–2171 (1999)
[10] Mortensen, P.P., Bro, R.: Real time monitoring and chemical profiling of a cultivation process. Chem. and Intell. Lab. Syst. 84(1-2), 106–113 (2005)
[11] Busscher, N., Kahl, J., Andersen, J., Huber, M., Mergardt, G., Doesburg, P., Paulsen, M., Ploeger, A.: Standardization of the biocrystallization method for carrot samples. Biological Agriculture and Horticulture 27, 1–23 (2010)