International Journal of Soft Computing and Artificial Intelligence, ISSN: 2321-404X, Volume-6, Issue-1, May-2018
http://iraj.in

A APPROACH IN FUZZY SETS FOR FEATURE REDUCTION

[1]LOKESH GOWDA J, [2]K V VISWANATHA

[1]Research Scholar, CMR University Bangalore
[2]Professor, CMR University Bangalore
E-mail: [1]lokeshgowda.ycm@gmail.com, [2]viswanathakv@yahoo.com

Abstract- Pattern representation refers to the number of classes, the number of available patterns, the number, type and scale of the features available to clustering algorithms. Pattern proximity is usually measured by a distance function defined on pairs of patterns. A variety of distance functions are in use in various communities. The Grouping step represents the organization of patterns into clusters based on pattern similarity. There are many clustering methods available, and each of them may give a different grouping of a dataset. The choice of a particular method will depend on the type of output desired, the known performance of method with particular types of data, the hardware and software facilities available and the size of the dataset [2].

Index terms- Pattern proximity,Clustering, Distance functions.

I. INTRODUCTION

The purpose of a learning algorithm is to determine which features are important for a particular decision as well as their relative importance. In order to find the value for D(x), the values for the two vectors w and x must be known. The values for x are obtained from the data. It is the job of the learning algorithm to determine the values for w. [1,5]. Supervised Learning: Data of known classification are used to determine important parameters (components of the feature vector) that contribute to the correct decision. Unsupervised learning: This algorithm estimates parameters of the statistical model of the classes from unlabeled training set in order to maximize an objective function.

II. BACKGROUND WORKS

In real life, the complete description of the classes is not known to provide partial information for optimal design of feature extractor or classifier. Often, in practice, one is called upon to solve problems involving sets of variables when it is known that there exists some inherent relationship among the variables. It may be of interest to develop a method of prediction. The concept of regression analysis deals with finding the best relationship between Y and x, quantifying the strength of that relationship, and the use of methods that allow for prediction of the response values given values of the regressor x. " Regression analysis" applies to situations in which relationships among variables are not deterministic i.e., not exact [16].

The ultimate goal of restoration techniques is to improve an image in some sense. The restoration can be viewed as a process that attempts to reconstruct or recover an image that has been degraded by using some a priori knowledge about the degradation phenomenon. Thus restoration techniques are oriented toward modeling the degradation and applying the inverse process in order to recover the original image. This approach usually involves formulating a criterion of goodness that will yield some optimal estimate of the desired result. The effectiveness of restoration techniques mainly depends on the accuracy of the image modeling.

Methods of random noise removal are to reconstruct the original (signal-only) image. They can be divided into two main groups: image filtration and imageaveraging [4, 28, 29]. In the first method image is either convolved with Gaussian mask or non-linearly filtered (for example with median filter). However, filtration affects with blurred appearance of an image and in result compromises the level of details.

One method of characterizing an object is by recording the observations related to the object, which constitute the feature values in a multidimensional space. Classification of multidimensional data set or cluster analysis is one of the Pattern Classification techniques and should be appreciated as such[11]. Feature selection is the process of identifying the most effective subset of original features to use in clustering. Feature extraction is the use of one or more transformations of the input features to produce new salient features. Either or both of these techniques can be used to obtain what is called a feature set (or feature vector)[9].

III. SIGNAL TO NOISE RATIO

Each digital image has two main components: a stable signal and a random noise in accordance with equation:

$$L' x, y) = L(x, y) + N(x, y) \qquad (1)$$

where: digital image (noisy image);
L- signal component (signal-only image);

N-noise component;
x, y-pixel coordinates.

Detection of the existence of secret message without the knowledge of embedding algorithm in binary noisy image is extremely difficult. The widely use of digital documents makes digital document image processing more and more useful. Data hiding in document images have received much attention recently and appears to be a new emerging technology. Some new techniques have been developed for data hiding in binary document images. One class of techniques for binary image data hiding is to change the value of individual selected pixels, such as the work in [13, 21, 22, 27].

Unification Of Embedding Process: To start with, we give a review of these embedding schemes by studying how these embedding algorithms preserve the quality of the images. These pixel-flipping techniques study the flippability of each pixel by comparing it with its neighboring pixels. In [21, 22], the authors come up with aflippablity score computation method by studying smoothness and connectivity in 3x3 neighborhoods. In [27], the authors choose 100 pairs of boundary patterns with the goal to preserve the overall shape of a character and minimize noticeable artifacts and distortion. In [13], the authors study the distortion introduced by flipping a pixel by subjective testing. In [14], the authors choose the pixels by flipping which the connectivity is preserved. Then these methods hide information by flipping the "flippable" pixels. One of the important observations is that most of the flippable pixels by these schemes can constitute an L-shape pattern with its 8-connected neighboring pixels. Normally, more than 80% pixels being flipped by these schemes satisfy this condition.

The reason that all these schemes choose the center pixel of the above mentioned patterns as flippable pixel is that flipping of the center pixel of the pattern will not affect the smoothness and connectivity of the image [17]. We define set A= {pixel (u, v) |the 3x3 window centered at (u, v) is an L-shape pattern}, where (u, v) represents the coordinate of a pixel. The pixel in set A is called a center of L-shape pattern pixel (COL pixel). In general, all these embedding schemes choose pixels from set A for flipping; of course, different schemes may use a different subset. Also, different scheme may choose some additional pixels, which do not belong to set A. In our analysis, we assume we have no knowledge in how the scheme chooses the subset from set A. As most of the pixels being flipped are pixels from set A, we detect the existence of secret message by detecting the flipping of pixels from set A.

Removing impulsive noises from scalar images is a problem of great interest since these short duration and high-energy noises can degrade the quality of digital images in a large variety of practical situations [10]. In this context of non-Gaussian noise, nonlinear processes are often invoked. Among these nonlinear processes median filtering is a classical tool leading to good results [6]. Nevertheless, these median filtering techniques involve strong statistics calculation and can turn out to be highly time consuming to compute. Another nonlinear process classically used for restoration tasks is the diffusion process of Perona-Malik [13]. This process, based on a variational approach, presents short implementation time and has the ability to remove noise while keeping edges stable on many scales.

IV. METHODOLGY

Several non-hierarchical methods have been proposed for clustering data. Among them, the k-means algorithm is perhaps the most popular one. It is a simple iterative hill-climbing algorithm where the solution obtained depends on the initial clustering. The histogram is perhaps the most common graphic for displaying the distribution of a single variable. While constructing a histogram seems to be straightforward, the appearance of the histogram is arbitrarily tied to the interval width. As an alternative to a histogram, statisticians have developed several smoothing algorithms to estimate the underlying shape of the data [23].

Cluster analysis is a formal study of algorithms and methods for grouping or classifying objects. A cluster analysis relieves a researcher of the treacherous job of looking at a pattern matrix or a similarity matrix to detect clusters, which has application even in divide and conquer strategy to reduce the computational complexity of various decision making algorithms in Pattern Recognition.
Distance measures are important issues in data exploration problems. A classifier relies on distance function between any two patterns. When it comes to classification of a set of samples, the determination of similarity and / or dissimilarity between the samples becomes the most important and fundamental step. A properly devised distance measure contributes a lot to the classification of the data.

The steps involved in the Implementation: Training process:
Step 1: Feature extraction of the preprocessed sample. The features under consideration in this context are duration, mean energy, maximum energy and maximum amplitude. These values are calculated and stored in a database.
Step 2: Classification: involves the following steps

Step 2.1: Mean is calculated for each feature of all the samples.

Step 2.2: Variance is calculated with respect to Mean for each sample.

Step 2.3: A Threshold value is calculated for the variance vector of theFeatures.

Step 2.4: In the Variance vectors, the values lesser than the threshold are

Labeled ‘0’ or Low, and the values greater than the threshold are Labeled ‘1’ or High . Since we have four features, we obtain a pattern string (Eg: H L H H). Hence, we obtained sixteen classes Based on different combinations of the pattern strings.

Step 3: The classified data is now stored in the knowledge base. This information is later used for clustering.

CONCLUSION

Estimation is based on a training set whose classification is known beforehand (e.g.: assigned by human experts). There are different distance measures available which deal with these problems with different concepts. But there is no single general distance measure, which can be applied on all types of data sets. If a classifier is efficiently designed it will be able to perform well on new patterns. In the preprocessing stage, we first identify the relevant features and then use a feature extractor to measure them.

Unsupervised learning can be used when little is known about the data set. It only requires a set of input vectors. Because of this lack of information, care must be taken in interpreting the output. First, if the number of categories in unknown, some means of deciding the appropriate number of clusters must be defined. One method is through the use of criterion functions. In any case, the results must be viewed in terms of the application to determine if they make sense and can be interpreted in terms of the problem.

REFERENCES

[1] A.K. Agrawala. (1970), Learning with a probabilistic teacher, IEEE Trans. Information Theory IT-16, pp 373-376.

[2] A.K. Jain, P.W. Duin Robert, J. Mao. (2000). Statistical Pattern Recognition: A Review; IEEE Transactions on pattern analysis and machine intelligence, 22, 1, 1-35.

[3] B. Burger, I. Ferrané, F. Lerasle, and G. Infantes, “Two-handed gesture recognition and fusion with speech to command a robot”, Autonomous Robots, vol. 32, issue 2, February 2012, pp. 129-147

[4] B. Jähne: “Digital image processing“, Springer, 1991.

[5] Bolles, Robert; Learning Theory; Holt , Rinchart and Winston, New York 1979

[6] Boncelet, C.G.: ‘Image noise models’ in ‘Bovik, A. (Ed.): ‘Handbook of image and video processing’ (Academic Press, New York, 2000)

[7] Catte, F., Coll, T., Lions, P., and Morel, J.: ‘Image selective smoothing and edge detection by nonlinear diffusion’, SIAM J. Appl. Math., 1992, 29, (1), pp. 182–193

[8] Cohen, N. Sebe, A. Garg, L. S. Chen, and T. S. Huang, “Facial expression recognition from video sequences: temporal and static modeling”, Computer Vision and Image Understanding, vol. 91, issues 1–2, July–August 2003, pp. 160-187.

[9] E. Dallas Johnshon, (1998). Applied Multivariate Methods for data analysis, Kanas University, Duxbury Press.

[10] E. DeRouin, J. Brown, L. Fausett, and M. Schneider, ``Neural Network Training on Unequally Represented Classes," in Intellligent Engineering Systems Through Artificial Neural Networks, (New York), pp. 135-141, ASME Press, 1991.

[11] E.A. Pattrick and F.P. Fischer. (1969), Clustering Mapping with experimental computer graphics. Symp. On Comm. Ploy.Inst.Brooklyn, April 8-10.

[12] F.-S. Chen, C.-M. Fu, and C.-L. Huang, “Hand gesture recognition using a real-time tracking method and hidden Markov models”, Image and Vision Computing, vol. 21, issue 8, 1 August 2003, pp. 745-758.

[13] G. Pan, Y.J.Hui, and Z. H. Wu, “ A novel data hiding method for two-color images”, in Lecture Notes in Computer Science . Oct. 2001, vol. 2229, Springer .

[14] H. Yang and A. C. Kot, "Data Hiding for Text Document Image Authentication by Connectivity-Preserving” in Proc. of IEEE ICASSP, pp II 505-508,March 2005

[15] Harmer, G.P., Davis, B.R., and Abbott, D.: ‘A review of stochastic resonance: circuits and measurement’, IEEE Trans. Instrum. Meas., 2002, 51, pp. 299–309

[16] Hosmer, David and Stanley Lemeshow (1989). Applied Logistic Regression. Y: Wiley & Sons. A much-cited recent treatment utilized in SPSS routines.

[17] J. Cheng and A. C. Kot, “Objective distortion measure for binary images”, in Proc. of IEEE TENCON 2004, pp355-358

[18] J. Onton, A. Delorme, S. Makeig, “Frontal midline EEG dynamics during working memory”, NeuroImage, Vol. 27, Issue 2, 15 August 2005, pp. 341-356.

[19] J. Stückler, et al., “Dynamaid: Towards a personal robot that helps with household chores”, In Robotics: science and systems conference (RSS’09), June 2009

[20] K. Kira and L.A. Randell. (1992), The feature Selection problem: Traditional Methods and a New Algorithm. Proc. of AAAI-92.

[21] M. Wu and B. Liu, “Data Hiding in Binary Image for Authentication and Annotation”, IEEE Trans. On Multimedia, vol. 6, NO. 4, pp528-538, August 2004

[22] M. Wu, E. Tang, and B. Liu, “Data hiding in digital binary image,” in IEEE ICME 2000. New York City, NY, USA, July 2000.

[23] M.E. Hellman. (1970), The nearest Neighbour classifiaction rule with a rejection option. IEEE Trans. Sys.Sci.Cyber. SSC-6.

[24] Maheshwari, O., and Ebenezer, D.: ‘Simultaneous removal of positive and negative impulses in images by using adaptive length recursive weighted median filter’, WSEAS Trans. Commun., 2005, 4, (12), pp. 1350–1355

[25] P. Ekman, “Strong evidence for universals in facial expressions: a reply to Russell’s mistaken critique”, Psychol. Bull., vol. 115(2), March 1994, pp. 268–287.

[26] Perona,P., and Malik, J.: ‘Scale-spaceand edge detection usinganistropic diffusion’, IEEE Trans. Pattern Anal. Mach. Intell., 1990, 12, (7), pp. 629–639

[27] Q. Mei, E.K. Wong, and N. Memon, “Data hiding in binary text documents,” SPIE Proc Security and Watermarking of Multimedia Contents III, Jan. 2001

[28] R. A. Schowengerdt: “Remote sensing models and methods for image processing”, Elsevier, 1997.

[29] R.C. Gonzalez, R. E. Woods: “Digital Image Processing (2nd Edition)”, Prentice Hall, 2002.

[30] T. Wang, J. Deng, and B. He, “Classifying EEG-based motor imagery tasks by means of time-frequency synthesized spatial patterns.” Clinical Neurophysiology, vol. 115, issue 12, December 2004, pp. 2744-2753.