

Mohamad Zairi Bin Abd Ghani, Anisah Binti Kamsin, Jeya Jassvine A/P Jeyabala Sundram
Faculty Information and Communication Technology,
University Technical Malaysia Melaka.
 {B031910029, B031910156, B031910136}@student.utem.edu.my

Summary On Five Technical Papers On Fuzzy Applications in Different Domains

1. FCM FACIAL EMOTION

In this research, fuzzy clustering techniques is used to determine category of facial reactions as well as the degree of expression. When the dimension is reduced, this clustering technique will also be able to choose the best description of faces.

The implementation of facial expression recognition and degree estimation involves 4 major steps. It begins with facial data acquisition. In this research, the train and test data has been collected from the Cohn-Kanade AU-Coded Facial Expression Database.

Next it is the Data preprocessing. The main challenge that will be faced is the brightness as well as the distribution of the facial images and geometric correspondence. To overcome this, an affine transformation is used to normalize the face geometries position and maintain face magnification and also to ensure gray values of each face has close geometric correspondence.

Next it is the facial data extraction process. Here we need to come in familiar terms with the Gabor wavelets. The Gabor wavelets are kernels that are similar to the 2D receptive field profiles of the mammalian cortical simple cells. Gabor wavelets are band pass filters that are used in image processing as for feature extraction and texture analysis. The Gabor wavelet can be defined as per this formula:

$$\psi_{\mu,\nu}(z) = \frac{\|k_{\mu,\nu}\|^2}{\sigma^2} e^{-\frac{\|k_{\mu,\nu}\|^2}{2\sigma^2}} \left[e^{ik_{\mu,\nu}z} - e^{-\frac{\sigma^2}{2}} \right]$$

Despite the sub sampling technique, the dimensionality was too large to classify hence a simple statistical method is used to reduce the dimensionality while minimizing mean squared reconstruction error.

Following is the clustering technique itself. FCM is targeted to find cluster centers or known as centroids to minimize dissimilarity function. To reach minimum level of dissimilarity, these equations can be used:

$$c_i = \frac{\sum_{j=1}^m u_{ij}^m x_j}{\sum_{j=1}^m u_{ij}^m}$$

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{kj}} \right)^{\frac{2}{m-1}}}$$

As for the results of FCM, there was a total of 945 facial images classified as neutral-happy sequences by 50 actors. Initially the data was divided into 2 groups to be used in the PCA whereby 200 faces are used to compute the covariance matrix and the remaining 745 are to be used to be projected into this covariance matrix's principal components. Out of the 745 images, 4 randomly selected subjects which equate to 54 faces are used to test the clustering approach. Here FCM is used to cluster the balance 691 images into 3 fuzzy clusters. One of the main observation in this subject is that the membership function is similar to the ideal trapezoidal fuzzy membership function and when the winner-take-all criteria is used to assign the absolute membership, the first person goes slowly to the maximum intensity on the other hand the condense individual remains longer in the maximum intensity.

The next observed results are from the neutral-surprise faces where similar steps have been applied but on 1173 images which have been played by 63 actors. In this subject, 209 faces were kept for PCA and remaining 964 faces were used for the clustering approach. The cluster center is calculated using the other 892 faces. From this test, it is figured that the best pair of principle components is chosen based on minimum distance criterion (MDC).

As a conclusion, it has been proven that fuzzy clustering is a good approach to estimate degree of intensity of facial expressions when a face is characterized with Gabor kernels and projected in two dimensional space and while using PCA. The best suitable principle is using MDA.

In my opinion, FCM is a decent technique for image classification because of its robust nature as well as it is conventional for clustering. Besides that, using Gabor kernels and projecting it into lower dimensional made it easier for the algorithm to detect any frequency content in the images as Gabor kernels are orientation sensitive filters that are widely used for texture analysis in image segmentations. In my opinion, the Gabor filter also represents human visual systems just like how humans can identify expressions and how happy or shocked someone is, the Gabor kernels function the same way.

2. FCM UNDERGROUND WATER FORECAST

Maintaining the stability of groundwater exploitation is a key issue in attaining efficient and sustainable water supplies. In this paper, we proposed a hybrid model that could effectively forecast groundwater levels and improve the efficiency of the process of their management. The hybrid model combined the fuzzy C-mean clustering algorithm (FCM) and singular spectrum analysis (SSA) It is able to define the range of possible total flow of the groundwater source, primarily in dry season periods. The model is related to the forecasting of the future states of the groundwater level (the general drawdown) using data obtained during the period of exploitation. The model is composed of two stages: in the first stage, we make fuzzy states of monitored data, while in the second, we forecast the future state. By using a fuzzy C-mean clustering algorithm, the original time series is divided into an adequate number of fuzzy states.

In this paper, we study the forecasting time of the invariant fuzzy time series of groundwater levels. Finally, the developed model produces the interval time series, characterized by the minimum and maximum value of the groundwater level for every point in the future. The developed model was tested by using the real data obtained by

monitoring the groundwater source Perminac. The groundwater source contains 14 wells with a maximum total capacity of 131 l/s and an average of 90 l/s. In recent years, overexploitation caused a significant decrease in the groundwater level in the wider area of the groundwater source.

Song and Chissom defined fuzzy relations among fuzzy time series, which are based on the assumption that the values of fuzzy time series $F(t)$ are fuzzy sets, and the observation of time t is caused by the observations of the previous times. In order to divide the observed data into an adequate number of fuzzy states, we apply the fuzzy C-mean clustering algorithm [20–23] over the set $S = \{s_i\}$, $i = 1, 2, \dots, N$. Every cluster is a fuzzy set defined by the relative closeness of space S . In this article, we look at the fuzzy time series of state to state relationships. We transform these into a deterministic forecasting model based on the singular spectrum analysis. This approach enables us to apply a detergent forecasting model for predicting future events in the future.

The process of the transformation of the fuzzy time series into a crisp time series is based on the fact that each fuzzy state $A_m(t)$, $m \in \{1, 2, \dots, M\}$ can be represented by a corresponding centre of the cluster. The first stage of the algorithm, called decomposition, includes the following two steps: embedding and singular value decomposition (SVD). The second stage of the algorithm, called reconstruction, includes the following two steps: grouping and diagonal averaging or Hankelization. The third stage of the algorithm concerns the future states of the groundwater level and is based on the linear recurrent formulae.

For me, the idea in this paper is very brilliant. As we know, Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. Groundwater levels play a critical role in basic processes connected to the efficiency and sustainability of water supply, hence the capacity to forecast them is quite important. The time-unbalanced water consumption rate and the worsening of the recharge conditions of captured aquifers have resulted in a significant deal of confusion in our understanding of how water supply systems work. By applying this hybrid model, it provides engineers a lot more leeway when it comes to controlling and planning long-term water supplies. It is also a

good improvement since the main aim of this paper was to develop a forecasting model for future states of the groundwater level (the general drawdown) using data obtained during the period of exploitation.

3. FCM METOBOLOMICS

Fuzzy c-means clustering (FCM) is an unsupervised method derived from fuzzy logic that is useful to solve multi-class and ambiguous clustering problems. FCM is directly applied on data matrix to generate a membership matrix which represents the degree of association on the samples between each clusters and the fuzziness coefficient -which denotes the degree of fuzziness in the algorithm. Data of metabolic profiles from 3 gene types (*Escherichia coli*) were used based on their metabolic profiles. For the sample preparation, three gene types of *E. coli* which is DC100 - mutant of deletion of SDH gene), DC101 mutant deletion acetate kinase and *E. coli* is the wild type to demonstrate the methods.

As for the data processing, firstly CSV files were converted into 2D matrices and exported as binary files. Then peak alignment between different chromatograms were performed. Then the peak tables generated was merged into one data matrix. Then the total ion intensities (TIC) of each chromatogram value is calculated and peak volumes were normalized to the respected TIC values. After all these processes, the final matrix contained 27 samples and 888 metabolites. Since the samples has great range of concentration, multivariate analysis has to be performed. The variables in the data matrix were scaled by range. After scaling each variable was between 0 and 1 which reduces teh influence of concentration range.

Next, PCA was applied but it did not have enough ability for the data hence PLS-DA was applied. From this, the wild type was further away from the other two. Relative distances between the groups in the score plot indicated the relative degree of metabolic differences. This basically indicates the wild type is different compared to the other two. There are also possibilities of overfitting, hence, in this research permutation test has been carried out to observe any overfitting. It is found that even when this model, PLS-DA was

good in terms of cross validation results and cumulative values, it is still not entirely reliable.

As for the results, the final metabolite matrix contained 27 samples and 888 metabolites, Since the metabolites in the samples had a great range of concentration, the influence of concentration in multivariate analysis had to be taken in consideration. Hence, the step taken was to range scaling before applying PCA, PLS and PLS-DA. PCA did not have enough ability to model the data hence PLS-DA was used. Three gene types of *E. coli* was differentiated in the score plot which indicates the relative degree of metabolic differences. Next a permutation test was carried out to monitor the overfitting.

FCM was then used to model data and classify the samples on the basis of metabolic profiles, here cross-validation was used to test overfitting. The membership matrix calculated by FCM was used as predictive matrix in PLS. After changing the settings of values of *C* and *m*, it can be seen that the cumulative values increased rapidly. This means that FCM can displace a better data fitting with higher degree of fuzziness. However, after investigating the membership values, it is learned that the membership values of two of the clusters were identical which means they completely overlap. The algorithm could still converge to fit the data but then provides poor predictive performance. So from the FCM classification, it is learned that that Wild type belonged to class 1 and gene types DC101 and DC100 belonged to class 2 and cannot be distinguished.

The application of FCM in discovering significantly changed metabolites is the fact that compared with the wild type, many metabolites in the gene knockout types have been changed, therefore it is important to discover the most significantly changed which can be done by combining FCM with PLS.

In my opinion, FCM is flexible and robust especially when dealing with ambiguous grouping and multiclass. Together with PLS, they can provide a good and powerful research method not just for metabolomics but other trends as well by providing a conclusion of improved visualization, classification as well as outlier estimation. I also learned from this process that analysis and visualization is extremely crucial to come to a conclusion and correlation to find a pattern.

4. FCM IMAGE WATERMAKING

Digital watermarking is a suggested process to protect copyright, verification of data integrity and etc. Most existing fractal based watermarking techniques have been developed according to the classification of the fractal codes, search region is divided into two halves and the watermark is concealed with regards to which half the best-pair domain block belongs to. A binary watermarking technique based on classification of fractal codes has been proposed. In this method, a part of original maps is replaced with modified ones. As a result, a binary watermark is placed into the image. In this paper, a new method for watermarking is presented which utilizes the fuzzy C-mean clustering (FCM) to address the watermark bits. The proposed method has high robustness against JPEG compression, Median Filter, and additive noise. To reduce the time that consumed for fractal image watermarking, a local search region is used in the proposed method.

A secret key is used to determine which P range blocks represent each bit of watermark. This secret key constructed by sender and should be known by receiver. Through watermark insertion, P range blocks are selected randomly for each bit and these range blocks should be saved in secret key in a safe manner. There are 5 step each in Insertion and Detection of watermark. For watermark insertion, firstly, the fractal encoding is applied on the original image to produce fractal codes for all range blocks. Secondly, for all domain blocks, the fuzzy C-mean clustering is used to classify then to four groups. Next, class A and B are determined according to the centers of the clusters which is calculated in previous step. Fourth step is, for each watermark, if the bit is “zero”, P range block that their matched domain blocks are belong to class A are selected randomly. But if the bit is “one”, P range blocks that their matched domain blocks are belong to class B are selected randomly. Last, watermarked image is constructed by fractal decoding process.

For Detection of watermark, firstly, fractal encoding is applied on watermarked image to produce fractal range blocks. For all domain blocks, the fuzzy C-mean clustering is used to classify them to four classes. Class A and B are determined according to the centers of the cluster which calculated in previous step. According to the secret key, for all P range blocks that determine a bit of watermark, if the

most matched (*Max Operator*) domain blocks are members of class A then the bit is “zero”. While if the most matched domain blocks are member of class B then the bit is “one”. If the most matched domain blocks are member of class C and D then the bit is “undetermined”. Step 4 is done for all bits of watermark according the secret key.

The parameters of the fractal affine transform are contrast scaling and range block mean. The watermark insertion procedure embeds binary watermark into an image according the values of horizontal and vertical derivatives of domain blocks instead of spatial positions of domain blocks that most of the previous similar works used. The Fuzzy C-Mean (FCM) clustering is used to partition domain blocks according to their vertical and horizontal

derivatives into four classes. FCM clustering is fast enough and doesn't affect the speed of watermarking procedure. In the future, we are going to use a global search by FCM to classify domain blocks and reduce the number of domain blocks to be checked. We can also use all four classes to represent a quad nary watermark and increase the capacity.

5. FCM BRAIN TUMOR

Tumours are of different types and they have different Characteristics and different treatment. As it is known, brain tumour is inherently serious and life-threatening because of its character in the limited space of the intracranial cavity (space formed inside the skull). To avoid that, this project uses computer aided method for segmentation (detection) of brain tumour based on the combination of two algorithms. In addition, it also reduces the time for analysis. At the end of the process the tumour is extracted from the MR image and its exact position and the shape also determined.

Normally the anatomy of the Brain can be viewed by the MRI scan or CT scan. The MRI scan is more comfortable than CT scan for diagnosis. Two algorithms are used for segmentation. So, it gives the accurate result for tumour segmentation. Tumour is due to the uncontrolled growth of the tissues in any part of the body. Normally brain tumour affects CSF (Cerebral Spinal Fluid). It causes for strokes. For the accurate detection of the malignant tumour that needs a 3-D representation of brain and 3-D analyzer tool. Image segmentation

plays a major role in the field of biomedical applications. Clustering is one of the widely used image segmentation techniques which classify patterns in such a way that samples of the same group are more similar to one another than samples belonging to different groups. Three-dimensional (3-D) processing and visualization of medical images is a rapidly growing area of research and MRI has provided a means for imaging tissue at very high resolutions providing the desired information for use in fields like reparative surgery, radiotherapy treatment planning, stereotactic neurosurgery, and others.

The existing method is based on the thresholding and region growing. The thresholding method was ignored the spatial characteristics. Normally spatial characteristics are important for the malignant tumour detection. In the thresholding based segmentation the image is considered as having only two values either black or white. But the bit map image contains 0 to 255 grayscale values. So sometimes it ignores the tumour cells also. In case of the region growing based segmentation it needs more user interaction for the selection of the seed. Seed is nothing but the centre of the tumour cells; it may cause intensity in homogeneity problem. And also it will not provide the acceptable result for all the images.

In proposed method, Magnetic resonance imaging (MRI) segmentation is used. It proposes a new method for estimating the right number of segments and automatic segmentation of human normal and abnormal MR brain images. The purpose of automatic diagnosis of the segments is to find the number of divided image areas of an image according to its entropy and with correctly diagnose of the segment of an image also increased the precision of segmentation. Segmentation is carried out by advanced K-means and Fuzzy C-means algorithms. Feature extraction is by thresholding and finally, Approximate reasoning method to recognize the tumour shape and position in MRI image using edge detection method. In the literature survey many algorithms were developed for segmentation.

K-Means is the one of the unsupervised learning algorithm for clusters. Clustering the image is grouping the pixels according to the same characteristics. In this paper input image is converted into Standard format 512 X 512, then find the total no. of pixels using Length =

Row X Column. Then convert 2D image into 1D and create no. of clusters depend on user. The k-means algorithm initially it has to define the number of clusters k.

Compute the cluster means m:

$$M = \sum_{i:c(i)=k} X_i / N_k$$

Then, calculate the distance between the cluster center to each pixel:

$$D(i) = \arg \min \|x_i - M_k\|^2, i = 1, \dots, N$$

Fuzzy C-Mean (FCM) is an unsupervised clustering algorithm that has been applied to wide range of problems involving feature analysis, clustering and classifier design. The fuzzy logic is a way to processing the data by giving the partial membership value to each pixel in the image which is ranges from 0 to 1. The clusters are formed according to the distance between data points and cluster centres are formed for each cluster. The Algorithm Fuzzy C-Means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. Fuzzy c-means (FCM) is the clustering algorithm which allows one piece of data may be member of more than one clusters. It is based on reducing the following function:

$$Y_m = \sum_{i=1}^N \sum_{j=1}^C M_{ij}^m \|x_i - C_j\|$$

Where,

m - any real number greater than 1,

M_{ij} - degree of membership of X_i in the cluster j,

x_i - data measured in d-dimensional,

R_j - d-dimension center of the cluster,

The update of membership M_{ij} and the cluster centres R_j are given by:

$$M_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{\|x_i - R_j\|}{\|x_i - R_k\|} \right)^{2/m-1}}$$

$$R_j = \sum_{i=1}^N x_i \cdot M_{ij}^m / \sum_{i=1}^N M_{ij}^m$$

The above process ends when,

$$\max_{ij} |M_{ij}^{(k+1)} - M_{ij}^{(k)}| < \delta$$

Where,

δ = termination value or constant between 0 and 1,

K= no of iteration steps.

C. The Fuzzy c-means Algorithm:

The algorithm contain following steps:

1. Initialize $M=[M_{ij}]$ matrix, $M(0)$

2. At k-step: calculate the centres vectors $R^{(k)}=[R_j]$ with $M^{(k)}$

$$R_j = \sum_{i=1}^N x_i \cdot M_{ij}^m / \sum_{i=1}^N M_{ij}^m$$

3. Update $U^{(k)}, U^{(k+1)}$

$$M_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{\|x_i - R_j\|}{\|x_i - R_k\|} \right)^{2/m-1}}$$

4. if $\|M^{(k+1)} - M^{(k)}\| < \delta$ then STOP, otherwise return to step 2.

As conclusion, there are different types of tumours are available. Suppose if it is a mass then K-means algorithm is enough to extract it from the brain cells. If there is any noise are present in the MR image it is removed before

the K-means process. The noise free image is given as an input to the k-means and tumour is extracted from the MRI image. And then segmentation using Fuzzy C means for accurate tumour shape extraction of malignant tumour and thresholding of output in feature extraction.

References

James, S.C.; Doherty, J.E.; Eddebbbarh, A. Practical Postcalibration Uncertainty Analysis: Yucca Mountain, Nevada. Groundwater **2009**, 47, 851–869. [[CrossRef](#)] [[PubMed](#)]

http://vasc.ri.cmu.edu/idb/html/face/facial_expression/, 15 May 2004.

B.Wohlberg and G.Jager, “A Review of the Fractal Image Coding Literature,” IEEE Trans. on IP, vol. 8, pp.1716-1729,1999

H.P. Ng, S.H. Ong, K.W.C. Foong, P.S. Goh and W.L. Nowinski. 2006. Medical Image Segmentation using K-Means clustering and Improved Watershed Algorithm. IEEE 2006.