A Fuzzy Expert System for Facial Expression Recognition

Masoumeh Rezaei[1], Mansoureh Rezaei[2]
[1]Department of Computer Engineering, Faculty of Electrical and Computer Engineering, University of Sistan and Baluchestan, Zahedan, Iran
[2]Department of Computer Engineering, Faculty of Electrical and Computer Engineering, University of Shiraz, Shiraz, Iran
mrezaei@ece.usb.ac.ir

***Abstract*—This paper presents a method for facial expression recognition using fuzzy expert system. The proposed expert system consists of two main steps: First, the pre-processing part, the feature extraction step provides sufficient information for the inference engine. For this reason, NMF is used to preserve the representation of the original image. Additionally, it guarantees that both of the resulting low-dimensional basis and its accompanying weights are non-negative. Second, it allows for creating rules with the SGERD algorithm and inferencing them. The second step applies a suitable set of fuzzy rules and aggregates them towards the final decision. We applied our approach to the Japanese Female Facial Expression dataset for recognizing the facial expression states. Experimental results demonstrate superiority of the proposed approach to the compared methods in terms of classification rate.**

***Index Terms*—Nonnegative Matrix Factorization, Fuzzy Expert System, SGERD algorithm.**

I. INTRODUCTION

Expert systems (ES), a branch of Artificial Intelligence (AI), means a comprehensive set of task-specific knowledge which is transferred from human to computer. The basic idea of ES is expertise. Fuzzy Expert System has been introduced to improve ES [1]. It is a kind of expert system that uses a collection of fuzzy membership functions and rules, instead of Boolean logic. Membership functions and "If… Then…"rules are used by Fuzzy Expert System for the expert's reasoning process.

Fuzzy set theory has been performed to model relevant features, while fuzzy logic has been utilized to make decision. Fuzzy Expert System compresses the knowledge base and reduces the complexity of the system using uncertain information. Thus, fuzzy expert system has been used in many real applications [13, 14, 28].

Researchers claimed that 7 percent and 38 percent of information is passed by language and tune respectively. However, 55 percent of information is transferred by verbal expression during human interaction. These data clearly indicate that facial expressions play an important role in human communications [3]. To design a facial expression recognition system, a standard dataset should be firstly provided. Then, discriminative component-based features are extracted from the images. These features should be quantitatively described by the facial parts such as lips, eyes and eyebrows, and flexion or extension of the facial muscles. Researchers have conducted studies to highlight the problems of facial expression recognition. These studies contain methods that use Local Binary Patterns (LBP) [19], static topographic modelling [21], texture and shape information fusion [12], Principal Component Analysis (PCA) [3], appearance modelling [1] and integration of facial expression with facial appearance models [20]. In the current study, we present a method for facial expression recognition using the Steady-State Genetic Algorithm for Extracting Fuzzy Classification Rules from Data (SGERD) [17] expert system. For this purpose, Japanese Female Facial Expression (JAFFE) dataset was used. Non-negative Matrix Factorization (NMF) was used to extract features and SGERD algorithm was performed create the rules. The rest of this paper is organized as follows: In Section 2, NMF and SGERD algorithm is described respectively. Then, the proposed method is explained in detail. Section 3, presents the experimental results. Section 4 describes the experimental part of the paper and last section concludes the paper.

II. METHODS

In this section, NMF algorithm is firstly explained. Then, SGERD algorithm is briefly introduced and the proposed algorithm is described.

A. Non-Negative Matrix Factorization

NMF is a low-rank approximation technique for unsupervised multivariate data decomposition. It is similar to Principal Component Analysis (PCA) and Independent Component Analysis (ICA), but they differs from each other with respect to their constraints and interpretations. NMF was firstly introduced by Lee and Seung in 1999. Many studies have been applied NMF algorithm for extensions and applications in image processing [8, 9, 10, 18, 22-27], signal processing [4, 5], and data mining [2, 15]. NMF efforts to decompose a given non-negative data matrix (e.g. Image, document) $A \in R^{m \times n}$ into a multiplication of two non-negative matrices $W \in R^{m\times k}$ and $H \in R^{k\times n}$ so that these matrices minimize the following criterion:

$$f(W,H) = \frac{1}{2}\|A - WH\|_F^2 \qquad (1)$$

Where k< min (m,n) is a positive integer that specifies the rank of NMF and F is the Frobenius norm. For the case of facial expression recognition, n is the number of images and each column of A is an image with size $w\times h$. In other words, the dimension of m must be $w \times h$ so that the original number of the features is m and reduced to k by NMF decomposition.

In general, NMF algorithms are divided into three general classes: multiplicative algorithms, gradient descent algorithms, and alternating least square algorithms [2]. These updating algorithms attempt to find more suitable values for the W and H matrices, iteratively. These are terminated when the approximate equality of A≈WH with an acceptable error is satisfied. In the present study, we used the first form of these three categories. The original algorithm is started by randomly initializing the matrices W and H with non-negative values. Then, it is iteratively updated according to the following equations:

$$H = H.*\,(W^T A)./(W^T WH + 10^{-9}) \qquad (2)$$

$$W = W.*\,(AH^T)./(WHH^T + 10^{-9}) \qquad (3)$$

B. SGERD Algorithm

SGERD is a steady-state genetic algorithm to extract fuzzy classification rules. Generally, the classification rules are utilized based on following form:

$$Rule: if\ X_1\ is\ A_{j1} \ldots X_n\ is\ A_{jn}\ then\ Class\ C_j \quad for\ \ j = 1,\ldots,N \qquad (4)$$

Where $X = [x_1, x_2, \ldots, x_n]$ and $A_{ji}(i = 1,2,\ldots,n)$ are n-dimensional pattern vector and antecedent linguistic value respectively. C_j is the consequent class of R_j while N is the number of rules.

In SGERD algorithm, each chromosome of the population is a fuzzy rule. Additionally, every antecedent variable that takes one fuzzy set from $\{L0, L1, .., L14\}$ is a gene (Figure 1). The fuzzy set $L0$ is 'don't care' and its membership value is always 1. To initialize the population, all fuzzy rules which have only one active antecedent variable are selected. Each variable can take one of the 14 fuzzy sets; so, the number of population will be 14 × n at most. Then, their consequent class is determined and these rules are grouped into M classes. After selecting the best Q per class, the initial main population will be M × Q rules as a maximum set. To determine the fitness of a rule, a support measure is used as follow:

$$SUP(A_j \Rightarrow Class\ T) = \frac{(\sum X_{p\in Class\ T}\, \mu_j(X_p))^2}{\sum_{p=1}^{m} \mu_j(X_p)} \qquad (5)$$

Where the compatibility grade of the training pattern X is determined with the antecedent part of rule Rj by the product operator as follow:

$$\mu_j(X) = \prod_{i=1}^{n} M_{ji}(X_i) \qquad (6)$$

Where $\mu_{ij}(.)$ is the membership function of the antecedent fuzzy set Aji and $Aji \in L0, L1, \ldots, L14$. To reproduce a child, each parent rule Rj randomly chooses another parent with similar consequent class. Then, one of its active antecedent variables is selected as a random (e.g., xi). If xi is inactive in Rj, it is activated by replacing $L0$ with one fuzzy set from $L1, L2, \ldots, L14$ and subsequently, all possible offspring (at most 14) are generated. The consequent class of offspring should be determined after reproduction. Since in each iteration one antecedent is activated, the number of generation is n.

Algorithm is organized as follow [17]:

Inputs: n labeled patterns of a k-dimensional M-class problem and Q.

Outputs: Possibly $R = M\times Q$ fuzzy classification rules.

1) $i = 1$ *(i: generation number).*

2) Generate all fuzzy rules having only one active antecedent variable.

3) Determine the consequent class of each candidate rule.

4) Divide the candidate rules into M groups according to their consequent class.

5) Rank, in descending order of their fitness values, the candidate rules in each group.

6) Choose the best Q rules from each class (i.e., possibly

R = M × Q rules in total) as the population in the i-th generation. Merely, in the first generation, choose the second best R rules as the auxiliary population and put away for mutation.

7) Increment i, if i > n, go to step 11.

8) Use all individuals in the previous generation (i.e., R rules) as parents and do reproduction (i.e., crossover, mutation, or elitism) on them. That is, for each parent rule, generate as offspring all fuzzy rules having one more active antecedent variable than its parent, provided each new offspring is fitter than its parent. In this case, the number of offspring will totally be R × 14 at most.

9) If no offspring fitter than the parent is produced in step 8, go to step 11.

10) Consider both parents and offspring in step 8 as candidate rules (at most C = R + R × 14 rules in total) and go to step 3.

11) Use R = M × Q rules (obtained in step 6 for the i-th generation) as the final population and stop. The actual length of these rules is i or less.

C. The Proposed Method

In the current study, we present a system that combines a structural and statistical method for feature extraction and classification technique based on fuzzy logic. NMF algorithm was used for feature extraction. It has the following benefits in comparison to the other feature extraction methods such as PCA in image processing field. Firstly, the elements of W are non-negative; consequently, basis columns can be visualized (Fig, 3). Secondly, non-negative elements of H make a non-subtractive combination of basis components. This property gives NMF part-based representation in contrast to other whole-based representation methods. Therefore, it is desirable in facial expression problem since only a part of an image is processed and facial expression is shown by a few parts of a face. Thirdly, NMF causes a sparse representation of input data in terms of W and H matrices lead to the drastic reduction in data storage [12].

After extracting features using NMF, rules should be created. For this purpose, each attribute was firstly rescaled to unit interval [0, 1], the pattern space was partitioned into fuzzy subspaces. In comparison to SGRED that used 14 fuzzy sets on each attribute, 5 fuzzy sets were applied in this study. Five triangular-shaped membership functions were considered for each feature and zero element for 'don't care'. A triangular function was defined by a lower limit a, an upper limit b, and a value m, a < m < b. These variables determine the x coordinates of the three corners of the underlying triangular MF.

$$\mu_A(x) = \begin{cases} 0 & x \le a \\ \dfrac{x-a}{m-a} & a < x \le m \\ \dfrac{b-a}{b-m} & m < x < b \\ 0 & x \ge b \end{cases} \quad (7)$$

The final MFs for each feature are shown in Figure 1.

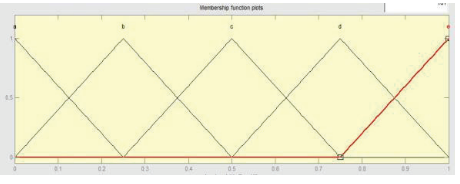

Figure 1: Membership function used for SGERD method.

Then, we used SGERD algorithm to extract the rules. For facial expression, seven classes are considered including anger, disgust, fear, happy, neutral and sad, making 7 variables for M.

For the reasoning step, the single winner reasoning method was performed. In this method, a new pattern $Xt = [xt1, xt2, \ldots, xtm]$ is classified according to the consequent class of winner rule R_w. On the other hand, the winner rule has the maximum compatibility degree with Xt among the fired rules. This can be viewed as follow:

$$\mu_w(X_t) = \max\{\mu_j(X_t)\} \quad j = 1,2,\ldots..,k \qquad (8)$$

The pseudo-code of the proposed algorithm is as follow:

Inputs: *Data matrix that is achieved by pre-processing phase.*

Output: *classification rate*

i=1;

c=1;

Repeat until all samples are evaluated

Set i^th column of Data matrix as test sample and the remaining as training samples(A matrix).

Set random matrices as W and H matrix.

For i=1: maximum iteration for NMF

Update H and W matrices trough equations (2) and (3)

End

Use H matrix as features of training sample.

Use SGERD algorithm for achieving proper rules.

*Set inverse(W)*old test features to achieve new test features.*

Determine the label of each pattern using equation (8)

If predicted label is equal to real label

Increment c;

Increment i;

Return c/number of samples.

III. RESULTS AND DISCUSSION

In our experiments, facial expression images of JAFFE dataset1 were used. This dataset consists of 213 images of 7 facial expressions. These expressions contain 6 basic facial expressions and a neutral expression posed by 10 Japanese females. At first, in the pre-processing phase, all faces from the images were cut and aligned into a fixed size (33×33) (Fig. 2). For each image, histogram equalization was applied and pixel's intensity of each matrix was normalized through [0, 1].

Figure 2: original image is illustrated in the left and image after preprocessing phase is demonstrated at right

After pre-processing phase, NMF was used to extract the features. Based on trial and error, a rank of NMF was assigned to 16 features. The features accomplished by NMF with 100 iteration are shown in Figure 2.

Figure 3: Features that are achieved with NMF

After that, the rules were extracted with SGERD algorithm and the output class was determined by the winner rule. At the final stage, a recognition rate at 88% was achieved in this study. The proposed method was compared to the K-Nearest Neighbors (KNN) method with k=1 and k=3 and Support Vector Machine (SVM) method. The results are shown in Table 1. According to our findings, our proposed method outperformed the other mentioned methods.

Table 1: Comparison of the proposed method, KNN and SVM in term of classification rate

Model	Classification rate	Execution time(s)
The proposed method	88.73%	24.091868
1-NN	82.15%	3.267575
3-NN	82.15%	3.202147
SVM	84.0%	5.433813

IV. CONCLUSION

The proposed method contains structural and statistical methods to extract features and a modeling and classification technique based on fuzzy logic. We used NMF algorithm for feature extraction. It is better than the other feature extraction methods with respect to its representation and reduction in storage. Then, we used SGERD algorithm to extract the rules and single winner method for reasoning. SGERD is one of the state-of-the-art algorithms for extracting the rules. It is used for its simplicity, fast and intuitive. Its generated rules are short, accurate, and interpretable. It can be applied to high-dimensional problems; therefore, it is suitable for image data. The obtained results illustrate that the proposed method outperformed the other methods in term of classification rate.

REFERENCES

[1] A.. Lotfizadeh, "Fuzzy sets. Information and Control". 8: 338–353. (1965).

[2] M.W. Berry, M. Browne, A.N Langville, V. Pauca, R. Plemmons , "Algorithms and Applications for Approximate Nonnegative Matrix Factorization". Computational statistics and data analysis. 52(1), 155-173, (2006)..

[3] A.J. Calder, A.M. Burton, P. Miller, A.W. Young, S. Akamatsu, "A principal component analysis of facial expressions". Vision Research. 41(9), 1179-1208, (2001).

[4] Z. Chen, A. Cichocki, "Nonnegative Matrix Factorization with temporal smoothness and/or spatial decorrelation constraints" ,Technical Report, Laboratory for Advanced Brain Signal Processing, RIKEN, Tokyo, Japan, (2005).

[5] A. Cichocki, ,R. Zdunek, S. Amari, "Csiszar's divergences for Non-negative Matrix Factorization: Family of new algorithms". Independent Component Analysis and Blind Signal Separation, 3889, 32 – 39, (2006).

[6] O. Cordon, M. J. Del Jesus, F. Herrera, "Aproposal on reasoning methods in fuzzy rule-based classification systems", Int. J. Approx. Reason., (20), 21–45, (1999).

[7] H. Ebine, O. Nakamura, "The recognition of facial expression based on fuzzy expert system". Electrical and Computer Engineering, (2), 565-568, (1998).

[8] T. Feng, S. Z. Li, H.Y.Shum, H. Zhang, "Local Non-Negative Matrix Factorization as a Visual Representation". Development and Learning, 178-183, (2002).

[9] L. Finesso, L. Spreij, "Nonnegative matrix factorization and I-divergence alternating minimization". Linear Algebra and its Applications, 416(2-3), 270-287, (2006).

[10] B. K. Jeon, Y.B. Jung, K.S. Hong, "Image segmentation by unsupervised sparse clustering". Application of computer vision, 2-7. (2005).

[11] M.Z. khedher, G. Ali-Talib, "A fuzzy expert system for recognition of handwritten arabic sub-words", Signal processing and its application. 1-4, (2007).

[12] I. Kotsia, S. Zafeiriou, N. Nikolaidis, I. Pitas, "Texture and shape information fusion for facial expression and facial action unit recognition". Pattern Recognition, 41(30), 833-85, (2008).

[13] F. Olusegun Folorunso , O. A. Mustapha, "A fuzzy expert system to Trust-Based Access Control in crowdsourcing environments", Applied Computing and Informatics. In press, (2014).

[14] Akashdeep, K.S. Kahlon, "An embedded fuzzy expert system for adaptive WFQ scheduling of IEEE 802.16 networks", Expert Systems with Applications, 41(16): 7621–7629, (2014).

[15] C. T. Li, Y. Yuan, R. Wilson, "Unsupervised clustering of conditional random fields approach for clustering gene expression time series". Bioinformatics. 24(21), 2467-2473, (2008).

[16] B. Abboud, F. Davoine, M. Dang, "Facial expression recognition and synthesis based on an appearance model". Signal Processing. 19(8), 723-740, (2004).

[17] E.G. Mansouri, M.J. Zolghadri, S.D. Katebi, "SGERD: A Steady-State Genetic Algorithm for Extracting Fuzzy Classification Rules From Data", IEEE Transactions on Fuzzy Systems, 16: 1061-1071. (2008).

[18] J.S. Zhang, C.P.Wang, Y.Q. Yang, "Learning latent features by nonnegative matrix factorization combining similarity judgments", Neurocomputing, 155: 43-52, (2015).

[19] C. Shan, S. Gong, P.W. McOwan, "Facial expression recognition based on Local Binary Patterns: A comprehensive study". Image and Vision Computing. 27(6): 803-816, (2009).

[20] P. Tsai, L. Cao, T. Hintz, T. Jan, "A bi-modal face recognition framework integrating facial expression with facial appearance". Pattern Recognition Letters, 30(12): 1096-1109, (2009).

[21] J. Wang, , L. Yin, "Static topographic modelling for facial expression recognition and analysis". Computer Vision and Image Understanding. 108(1-2): 19-34, (2007).

[22] S. Wild, J. Curry, A. Dougherty, "Improving non-negative matrix factorizations through structured initialization". Pattern Recognition, 37(11):2217-2232, (2004).

[23] Y. Xue, Ch. Tong, Y. Chen, W. Chen, "Clustering-based initialization for non-negative matrix factorization". Applied Mathematics and Computation, 205(2): 525-536, (2008).

[24] Z. Ying, , G. Zhang, "Facial Expression Recognition Based on NMF and SVM". International Forum on Information Technology, 458 – 462, (2009).

[25] L. Zhao, G. Zhang,, X. Xu, "Facial Expression Recognition Based on PCA and NMF". Intelligent control and automation. 6826 – 6829, (2008).

[26] D. Lee, H. Seung. "Learning the Parts of Objects by Non-Negative Matrix Factorization". Nature 401, 788–791, (1999).

[27] A. Langville, C. Meyer, R. Albright., J. Cox, D. Duling, "Algorithms, initializations, and convergence for the nonnegative matrix factorization". Preprint, (2006).

[28] Z. Xu, K. Gao, T.M. Khoshgoftaar, N. Seliya, "System regression test planning with a fuzzy expert system", Information Sciences, 259: 532–543, (2014)