Face recognition using fuzzy kernel learning vector quantization

Z Rustam\(^1\) and Rika\(^1\)

\(^1\)Department of Mathematics, University of Indonesia, Indonesia

Abstract. In recent years, face recognition is widely used in various aspects as a form of technology advancement. Various studies were conducted to improve the accuracy of face recognition. In this research, Learning Vector Quantization and Fuzzy Kernel Learning Vector Quantization were used as a method of classification. The data used in this research was Labeled Face in The Wild-a (LFW-a). This database has no restrictions such as background, expression, position, and so on. Based on test results using LFW-a database, face recognition using LVQ method has highest accuracy at 89.33% and FKLVQ method has highest accuracy at 89.33% as well.

1. Introduction
Biometric is a term for measurement of physical and behavioural characteristic of humans. Physical characteristics include fingerprint, finger knuckle, iris, face, voice, etc., meanwhile behavioural characteristics include walking gait, handwriting, etc. Biometric information has been widely used for identification system in recent years due to its uniqueness in each individual. One of the most popular biometric identification system is face recognition. Compared to other biometric information, face image is easier and more flexible to acquire even when the person is unaware of being scanned.

Inspired by biological neurons, Artificial Neural Network (ANN) is a network of elements called neurons. A neuron can either inhibit or excite a signal. Input signals are inhibited or excited based on numerical weights associated with each connection to the artificial neuron [1]. In this paper, neural networks architecture namely Learning Vector Quantization (LVQ) was used to classify Labeled Faces in the Wild-a database (LFW-a). LFW-a is a database of face images designed for face recognition study without limitation such as background, expression, etc. This database already has a name label for each subject.

An LVQ network has an input layer, a competitive layer, and a linear layer. Competitive layer aims to classify input vectors to subclasses, while linear layer aims to combine subclasses into target classes. Both layers have one neuron for each class. If the competitive layer has \(m\) neuron, then the layer can learn up to \(m\) subclass which is then combined by a linear layer to form the \(n\) target class \((m \geq n)\) [2].

LVQ has attracted much attention because of its simple and efficient learning process. However, LVQ has some issues such it only updates the winner neuron for each input. Moreover, it uses
Euclidean distance so that LVQ is sensitive to noise and outlier. In this paper, fuzzy membership was used to solve the first problem. To solve the second problem, kernelized distance measure was used as a replacement for Euclidean distance [3]. Thus, fuzzy kernel learning vector quantization was used to classify LFW-a database as well.

2. Classification
Classification is the process of finding a model that describes the class or category of output data, therefore it can be used to predict the class of new inputs. Classification algorithm is divided into two stages namely training and testing. In the training stage, the classification algorithm builds a model or classifier based on the learning process of training data containing input data and classes or categories of each data. The testing stage aims to find out the accuracy of the classification algorithm. At this stage, classifier classifies the testing data into classes. The accuracy of the classifier is said to be the classifier capability to correctly predict the class of testing data [4].

2.1. Learning vector quantization
LVQ is one of Artificial Neural Network (ANN) architecture. ANN is a classification method based on the principle of biological neuron network. LVQ network has 3 layers: input layer, competitive layer and linear layer [2]. Competitive layer has a winner-take-all rule, in the sense that neurons in competitive layer compete each other to become winner neurons [3].

The training process at LVQ aims to determine the weight vectors of competitive neurons, thus obtaining the training data mapping to each class on the competitive layer correctly. For each input vector, \( \mathbf{x}_k \), the weight vector, \( \mathbf{w}_i \), is updated with the formula as follows:

\[
\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + \alpha(t) h_{ik}(\mathbf{x}_k - \mathbf{w}_i(t)),
\]

where \( \alpha(t) \) is learning rate (\( 0 \leq \alpha(t) \leq 1 \)) and \( h_{ik} \) is indicative function whose value is 1 when \( i \) is the winner neuron and 0 when \( i \) is not the winner neuron [3].

The winner neuron is selected based on Euclidean distance:

\[
d(\mathbf{x}_k, \mathbf{w}_i) = \| \mathbf{x}_k - \mathbf{w}_i \| = \left( \sum_{j=1}^{R} (x_{kj} - w_{ij})^2 \right)^{1/2},
\]

where \( R \) is the number of elements in the vector [2].

3. Fuzzy membership
Inspired by the Fuzzy C-Means (FCM) algorithm, the primary key of FCM is the use of the membership function found in fuzzy sets. Fuzzy C-Means Algorithm provides the identity data for each category using fuzzy membership [5]. The objective function of FCM can be formulated as follows [6]:

\[
J_m = \sum_{i=1}^{c} \sum_{k=1}^{N} (u_{ik})^m \| \mathbf{x}_k - \mathbf{v}_i \|^2,
\]

where:

- \( c \) = number of clusters (\( c \geq 2 \))
- \( N \) = number of data
- \( m \) = degree of fuzziness (\( m \geq 1 \))
- \( \| \mathbf{x}_k - \mathbf{v}_i \| \) = Euclidean distance between \( \mathbf{x}_k \) and \( \mathbf{v}_i \)
- \( X = \{ \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \} \) is the input data set
- \( V = \{ \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_c \} \) is the cluster centre set
- \( U = \{ u_{ik} \} \) is the fuzzy membership.
The objective function has constraints [6]:

\[ 0 \leq u_{ik} \leq 1 \]
\[ \sum_{i=1}^{N} u_{ik} = 1 \quad \forall k = 1, \ldots, N \]
\[ 0 \leq \sum_{k=1}^{N} u_{ik} \quad \forall i = 1, \ldots, c \]

The membership function, \( u_{ik} \), was obtained as follows [6]:

\[ u_{ik} = \left( \sum_{j=1}^{c} \left( \frac{\| x_k - v_j \|}{\| x_k - v_i \|} \right)^{2(m-1)} \right)^{-1} \quad (3.2) \]

In LVQ algorithm, \( c \) denotes number of competitive neurons and \( v_j \) becomes \( w_j \) which is weight vector. The equation of \( u_{ik} \) membership becomes as follows:

\[ u_{ik} = \left( \sum_{j=1}^{c} \left( \frac{\| x_k - w_j \|}{\| x_k - w_i \|} \right)^{2(m-1)} \right)^{-1} \quad (3.3) \]

where:
\( c \) = number of competitive neurons \( (c \geq 2) \)
\( X = \{ x_1, x_2, \ldots, x_n \} \) is the input data set
\( W = \{ w_1, w_2, \ldots, w_c \} \) is the weight vector set
\( m \) = degree of fuzziness \( (m \geq 1) \)

4. Kernel

Let the set of input, \( X \), and nonlinear mapping functions, \( \phi : x \rightarrow \phi(x) \in F \), which maps \( x \) in the input space \( X \) to the feature space \( F \) which has a higher dimension. The kernel function is defined as the inner product of space \( F \) with

\[ K(x, y) = \phi(x)^T \phi(y) \]

for \( x, y \) in the \( X \) input space. With kernel \( K \), kernelized distance function is found as follows [7]:

\[ d(x, y) = \| \phi(x) - \phi(y) \| = (K(x, x) - 2K(x, y) + K(y, y))^{1/2}. \quad (4.1) \]

The kernel used in this paper was radial basis function kernel

\[ K(x, y) = \exp \left( \frac{\| x - y \|^2}{2\sigma^2} \right) \quad (4.2) \]

where \( \sigma \) is the parameter in the kernel function.

5. Fuzzy kernel learning vector quantization

The fuzzy kernel algorithm was constructed with objective functions as follows [6]:

\[ J_m = \sum_{i=1}^{c} \sum_{k=1}^{N} (u_{ik})^m \| \phi(x_k) - \phi(w_i) \|^2 \quad (5.1) \]

where:
\( c \) = number of competitive neurons \( (c \geq 2) \)
\( N \) = number of data \\
\( m \) = degree of fuzziness (\( m \geq 1 \)) \\
\( \| \phi(x_i) - \phi(w_j) \| = \) kernelized distance between \( x_i \) data and weight vector \( w_j \) \\
\( X = \{ x_1, x_2, \ldots, x_n \} \) is the input data set \\
\( W = \{ w_1, w_2, \ldots, w_c \} \) is the weight vector set \\
\( U = [ u_{ik} ] \) is the fuzzy membership. \\
The objective function has constraints: 
\[
0 \leq u_{ik} \leq 1 \\
\sum_{i=1}^{c} u_{ik} = 1 \quad \forall k = 1, \ldots, N \\
0 \leq \sum_{k=1}^{N} u_{ik} \quad \forall i = 1, \ldots, c 
\]
By simplifying \( \| \phi(x_i) - \phi(w_j) \|^2 \), we get the following equation [8]:
\[
\| \phi(x_i) - \phi(w_j) \|^2 = (\phi(x_i) - \phi(w_j))^T (\phi(x_i) - \phi(w_j)) \\
= \phi^T(x_i)\phi(x_i) - \phi^T(x_i)\phi(w_j) - \phi^T(w_j)\phi(x_i) + \phi^T(w_j)\phi(w_j) \\
= K(x_k, x_i) - K(x_k, w_j) - K(w_j, x_i) + K(w_j, w_j) \\
= 2 - 2K(x_k, w_j) 
\] (5.2)
By substituting equation (5.2) into equation (5.1), the objective function was obtained as follows [9]:
\[
J_m = 2 \sum_{j=1}^{c} \sum_{k=1}^{N} \left( u_{ik} \right)^m \left( 1 - K(x_k, w_j) \right) 
\] (5.3)
The kernelized membership function, \( u_{ik} \), was obtained as follows:
\[
u_{ik} = \left[ \frac{\sum_{j=1}^{c} \left( 1 - K(x_k, w_j) \right)}{\sum_{j=1}^{c} \left( 1 - K(x_k, w_j) \right) \frac{1}{m-1}} \right]^{1/(m-1)} 
\] (5.4)
Next, find the minimum distance between input vector, \( x_k \), and weight vector, \( w_j \), as follows
\[
d^2(x_k, w_j) = K(x_k, x_k) - 2K(x_k, w_j) + K(w_j, w_j) 
\] (5.5)
By simplifying the formula (5.5) using gradient descent, the equation \( w_j(t+1) \) can be formed as
\[
w_j(t+1) = w_j(t) - \alpha(t) h_{ik}(t) \frac{\partial d^2(x_k, w_j)}{\partial w_i} \\
w_j(t+1) = w_j(t) - \alpha(t) h_{ik}(t) \frac{2}{\sigma^2} K(x_k, w_j(t)) (x_k - w_j(t)) 
\] (5.6)
and \( h_{ik}(t) \) was calculated as follows:
\[
h_{ik}(t) = \left[ \frac{u_{ik}}{\max_{1 \leq i \leq c} \{ u_{ik} \}} \right]^{1/\sigma^2} 
\]
6. Experimental result
The database used in this paper was LFW-a database. It consists of 30 subjects with each subject having 10 face images. This database was designed for studying the problem of unconstrained face recognition. All face images in this database have different backgrounds, expressions, and face positions. Some subjects in the database also wear glasses and hats. Each face image in this database has been labelled according to the person's name in the image. The face image in this database was a grayscale image with .jpg format.

A digital image can be defined as a two-dimensional function, \( f(x, y) \), where \( x \) and \( y \) are the plane coordinates and the value of \( f(x, y) \) is the intensity or gray level of the image at that point. A digital image, \( f(x, y) \), sized \( m \times n \) pixels can be represented as a matrix containing \( m \) rows and \( n \) columns. A row vector can also be a representation of an image. The first row of \( f(x, y) \) becomes \( n \) first element of the vector, the second row becomes the next \( n \) element, and so on to form a \( 1 \times mn \) row vector [10]. Each face image was converted into a row vector to be used as an input vector on the algorithm. The LFW-a database was divided into two parts: training data and testing data.

6.1. LVQ experiment
The experiment using Learning Vector Quantization method was performed to identify the name of subject in face images in LFW-a database. The experiment was done using 10% training data from the total number of databases, where 1 out of 10 face images of each subject is taken for training and 9 face images from each subject were used for testing. In addition, there are also experiments with 20% training data, 30% training data, 40% training data, 50% training data, 60% training data, 70% training data, 80% training data and 90% training data. The results of experiment with various number of training data as follows:

| Training Data | Accuracy | Running Time (s) |
|---------------|----------|-----------------|
| 10%           | 51.85%   | 12.72           |
| 20%           | 75.83%   | 12.05           |
| 30%           | 79.52%   | 11.95           |
| 40%           | 76.11%   | 16.95           |
| 50%           | 89.33%   | 16.86           |
| 60%           | 89.17%   | 24.02           |
| 70%           | 87.78%   | 26.40           |
| 80%           | 86.67%   | 48.84           |
| 90%           | 80.00%   | 50.77           |

As seen in Table 1, the highest accuracy obtained was 89.33% with 50% training data. The running time for Learning Vector Quantization is fast. For 50% training data, the running time is only 16.86 seconds.

6.2. FKLVQ experiment
The experiment was also conducted with Fuzzy Kernel Learning Vector Quantization classification method. Fuzzy Kernel Learning Vector Quantization method involves kernel parameter which in this case is the sigma (\( \sigma \)) of radial basis function kernel. The experiment was done using 50% training data,
so 5 images of each subject face were used for training and 5 remaining images were used for testing. Experiment was performed using different sigma values (0.0001; 0.001; 0.05; 0.1; 1; 5; 10; 50; 100; 1000). Here are the results obtained:

| Training Data | Sigma (σ) | Accuracy | Running Time (s) |
|---------------|-----------|----------|-----------------|
| 50%           | 0.0001    | 13.33%   | 1593.70         |
| 50%           | 0.001     | 78.00%   | 1628.14         |
| 50%           | 0.05      | 86.67%   | 1646.44         |
| 50%           | 0.1       | 89.33%   | 1612.30         |
| 50%           | 1         | 82.00%   | 1589.86         |
| 50%           | 5         | 78.00%   | 1592.72         |
| 50%           | 10        | 76.67%   | 1551.97         |
| 50%           | 50        | 76.67%   | 1559.38         |
| 50%           | 100       | 76.67%   | 1556.69         |
| 50%           | 1000      | 76.67%   | 1538.64         |

As seen in Table 2, the highest accuracy obtained was 89.33% with 50% training data and sigma 0.1. In this case, Fuzzy Kernel Learning Vector Quantization has longer running time than Learning Vector Quantization.

7. Conclusion
This paper studies face recognition to identify the identity of subject in LFW-a face image database. Learning Vector Quantization was used due to its simple and efficient learning process. To cover the weakness of this method such as its sensitivity to noise and outlier, fuzzy and kernel are implemented in the classification method. Based on the experiment using Learning Vector Quantization classification method to classify the LFW-a database, the accuracy reaches the maximum of 89.33% using 50% training data. In the experiment of Fuzzy Kernel Learning Vector Quantization classification method on LFW-a database, only 50% training were used but different sigma (σ) parameter values (0.0001; 0.001; 0.05; 0.1; 1; 5; 10; 50; 100; 1000) were implemented in the experiment. Accuracy reaches the maximum of 89.33% at the value of σ = 0.1.

8. References
[1] Engelbrecht A P 2008 Computational Intelligence: An Introduction (Chichester: John Wiley & Sons)
[2] Jun Y and Dian-Hong W 2005 Sequential face recognition based on LVQ networks Proc. Of 2005 IEEE Int. Workshop on VLSI Design and Video Technology pp 284-7
[3] Zhang D, Chen S and Zhou Z 2004 Fuzzy-kernel learning vector quantization Advances in Neural Networks-ISNN 2004 Lecture Notes in Computer Science pp 180-5
[4] Han J and Kamber M 2006 Data Mining: Concepts and Techniques (San Francisco: Morgan Kaufmann)
[5] Wulan A, Jannati M V, Rustam Z and Fauzan A A 2016 Application kernel modified fuzzy c means for gliomatosis cerebri 2016 12th Int. Conf. on Mathematics, Statistics, and Their Applications (ICMSA)
[6] Zanaty E 2012 Determining the number of clusters for kernelized fuzzy C-means algorithms for automatic medical image segmentation Egyptian Informatics J. 13 pp 39-58
[7] Lv H and Wang W 2005 Handwritten digit recognition with kernel-based LVQ classifier in input space Advances in Neural Networks-ISNN 2005 Lecture Notes in Computer Science pp 203-8

[8] Rustam Z and Frederica Y 2018 Insolvency prediction in insurance companies using support vector machines and fuzzy kernel c-means J. of Physics: Conf. Series 1028 012118

[9] Chiang H, Lee T and Lee J 2011 On-road speed sign recognition using fuzzy kernel-based learning vector quantization Proc. 2011 Int. Conf. on System Science and Engineering

[10] Gonzalez R C and Woods R E 2008 Digital Image Processing (New Jersey: Pearson Education)