Interval valued fuzzy sets $k$-nearest neighbors classifier for finger vein recognition

Nordiana Mukahar$^{1,2}$, Bakhtiar Affendi Rosdi$^1$

$^1$Intelligent Biometric Group, School of Electrical & Electronic Engineering, Engineering Campus, Universiti Sains Malaysia, 14300 Nibong Tebal, Pulau Pinang, Malaysia

$^2$Faculty of Electrical Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Malaysia

E-mail: nordi7410@tganu.uitm.edu.my, eebakhtiar@usm.my

Abstract. In nearest neighbor classification, fuzzy sets can be used to model the degree of membership of each instance to the classes of the problem. Although the fuzzy memberships can be set by analyzing local data around each instance, there may be still a lack of knowledge associated with the assignation of a single value to the membership. This is caused by the requirement of determining in advance two fixed parameters: $k$, in the definition of the initial membership values and $m$, in the computation of the votes of neighbors. Thus, the two fixed parameters only allow the flexibility of membership using a single value only. To overcome this drawback, a new approach of interval valued fuzzy sets $k$-nearest neighbors (IVFKNN) that incorporating interval valued fuzzy sets for computing the membership of instance is presented that allows membership values to be defined using a lower bound and an upper bound with the length of interval. The intervals concept is introduced to assign membership for each instance in training set and represents membership as an array of intervals. The intervals also considered the computation of the votes with the length of interval. In order to assess the classification performance of the IVFKNN classifier, it is compared with the competing classifiers, such as $k$-nearest neighbors (KNN) and fuzzy $k$-nearest neighbors (FKNN), in terms of the classification accuracy on publicly available Finger Vein USM (FV-USM) image database which was collected from 123 volunteers. The experimental results remark the strong performance of IVFKNN compared with the competing classifiers and show the best improvement in classification accuracy in all cases.

1. Introduction

The $k$-nearest neighbors (KNN) is one of the non-parametric classifier that has been extensively employed in various fields of pattern classification [1], machine learning, clustering [2] and various research areas in biometric. The basic principle of KNN introduced in [3] classifies the test sample by the $k$ training samples that are closest to the test sample within the neighborhood size of, $k$. The test sample is then assigned to the class that has majority votes in its $k$-nearest neighbors. Due to its simplicity and effectiveness, the KNN has received abundant of interest from researchers and many improvements has been proposed for achieving higher accuracy.

Beneath the several advantages offered by the KNN, there exist some drawbacks to be solved. The KNN classifier treats every training samples as nearest neighbors with equal importance regardless of its typicalness as a class prototype and its distance to the pattern to be classified [4]. The fuzzy $k$-nearest neighbors (FKNN) classifier overcome this problem in KNN by associating the degree of membership of each training sample with analyzing local data around each sample.
However, there exists an issue of lack of information associated with the assignation of single value to the memberships that is caused by the necessity of predefined two fixed parameter in advance [5]: a) initial membership values, \( k \), b) computation of the votes of the neighbors, \( m \). It indicates that the performance of FKNN classifier is sensitive and dependent on the selection of values considered for the \( k \) and \( m \) parameters. In order to overcome the problem of setting up fixed parameters, the interval valued fuzzy sets introduced in [6] is implemented in FKNN to allow membership values to be defined using interval bounds: a lower bound and an upper bound. Instead of using a single value of membership, the implementation interval valued fuzzy sets in \( k \)-nearest neighbors provides a greater degree of flexibility with the length of interval to the decision rule [5]. Furthermore, the implementation of intervals in this approach allows to explore different values for the \( k \) and \( m \) parameters to obtain different degrees of membership, and as a results improving the capabilities of interval valued fuzzy sets \( k \)-nearest neighbors (IVFKNN).

2. The interval valued fuzzy \( k \)-nearest neighbors classifier

This section provides the explanation of the IVFKNN classifier and it can be formally described as follows: let \( y \) be a testing sample and a set of training samples \( T = \{ x_j \in \mathbb{R}^d \}_{j=1}^N \) which belong to \( C \) classes. The training set consists of \( x_j \) as the training sample with \( N \) is the number of training samples in \( d \)-dimensional feature space.

Initially, for a new testing sample, \( y \), the IVFKNN find the \( k \)-nearest neighbors in training set, \( T \). The IVFKNN classifier measures the similarity between the testing sample and training samples in feature space using euclidean distance metric, which suit the most in classification task especially if the training data is normalized in the domain \([0,1]\) [5]. Next, each neighbor is assigned to fuzzy membership values in which the computation is performed using the intervals definition.

In IVFKNN, the interval values of fuzzy memberships is assigned to each neighbor using the following membership function that proposed in [4]

\[
U_c(x_j, k_{\text{var}}) = \begin{cases} 
0.51 + \left( \frac{nn_c}{k_{\text{var}}} \right) \times 0.49, & \text{if } c = \omega \\
0.51 + \left( \frac{nn_c}{k_{\text{var}}} \right) \times 0.49, & \text{otherwise} 
\end{cases} 
\]  

(1)

where \( nn_c \) are the number of neighbors belong to class \( c \) available among the \( k_{\text{var}} \) neighbors of the testing sample. Unlike the FKNN scheme which requires the parameter setting of \( k_{\text{var}} \) to be fixed in advance, the interval values for membership in IVFKNN use different values of \( k_{\text{var}} \). The \( k_{\text{var}} \) should not be set up to a lower value such as \( k_{\text{var}} = 1 \) or \( k_{\text{var}} = 2 \) because the local information of data will lost if very few neighbors are included in determining the degree of memberships [5]. The high value setting of \( k_{\text{var}} \) should also be avoided which resulting the memberships equal to the global distribution of classes in the training data [5]. For this configuration, the parameter setting for \( k_{\text{var}} \) is set to an integer value within the range of \([3, 21]\).

Instead of resulting an assignation of fuzzy memberships with a single value as FKNN scheme, the IVFKNN presenting the memberships as an array of intervals which consisting of lower and upper bound values as follows

\[
U_c(x_j) = [U_{\text{low}}(x_j), U_{\text{high}}(x_j)] 
\]  

(2)

The lower and upper bound values of fuzzy memberships are obtained as in equation 3 and 4 respectively.

\[
U_{\text{low}}(x_j) = \min[U_c(x_j, k_{\text{var}})] 
\]  

(3)

\[
U_{\text{high}}(x_j) = \max[U_c(x_j, k_{\text{var}})] 
\]  

(4)
The votes, \( V(x_j, c) \) cast by nearest neighbors is computed as in equation 5 at which the Euclidean norm, \( D(x_j) \) and degree of membership, \( U_c(x_j) \) for each neighbor are weighted to produce a final vote for decision rule.

\[
V(x_j, c) = U_c(x_j) \cdot D(x_j)
\]  

(5)

The degree of membership, \( U_c(x_j) \) for each neighbor is represented as an array of intervals as obtained using equation 2 [4]. The weighted Euclidean norm is defined as in equation 6 and it can be presented as an interval by selecting a possible range of values for \( m \) instead of single \( m \) in FKNN. The parameter \( m_{var} \) is assigned to an integer value \([1.1, 2.2]\) [5].

\[
D(x_j, m_{var}) = \frac{1/\|y - x_j\|^2/m_{var}-1}{\sum_{j=1}^{k} 1/\|y - x_j\|^2/m_{var}-1}
\]  

(6)

The weighted Euclidean norm in equation 6 can be represented as an interval as follows

\[
D(x_j) = [D_{low}(x_j), D_{high}(x_j)]
\]  

(7)

where the lower and upper bound values are obtained as in equation 8 and 9 respectively.

\[
D_{low}(x_j) = \min(D(x_j, m_{var}))
\]  

(8)

\[
D_{high}(x_j) = \max(D(x_j, m_{var}))
\]  

(9)

The degree of membership and the weighted Euclidean norm are in the interval form as in equation 10, the final products, \( I_j \) is computed as in equation 11.

\[
U_c(x_j) \cdot D(x_j) = [U_{low}(x_j), U_{high}(x_j)] \cdot [D_{low}(x_j), D_{high}(x_j)]
\]  

(10)

\[
I_j = [U_{low}(x_j) \cdot D_{low}(x_j), U_{low}(x_j) \cdot D_{high}(x_j), U_{high}(x_j) \cdot D_{low}(x_j), U_{high}(x_j) \cdot D_{high}(x_j)]
\]  

(11)

The final vote is then represented as an array of interval as follows

\[
U_c(x_j) \cdot D(x_j) = [\min(I_j), \max(I_j)]
\]  

(12)

Finally, the votes for every class is calculated as equation 13 and the testing sample is assigned to the class with the maximum vote overall as shown in equation 14. Since the votes is in the interval form, the Lexicographic procedure with respect to the lower bound is selected to provide an ordering on the intervals representing the votes assigned to each class. For example, let \( V(c_1) = [a, b] \) and \( V(c_2) = [c, d] \) denote the votes with interval values for class 1 and 2 respectively. The Lexicographic procedure provides an order between two interval votes, \([a \cdot b]\) and \([c \cdot d]\) by comparing \(a\) and \(c\) and choose the maximum value between them. If tie exist, then \(b\) and \(d\) are required to compare to break the tie and the class with maximum value is selected between those values [5].

\[
V(c) = \sum_{j=1}^{k} V(x_j, c)
\]  

(13)

\[
y = \arg \max \{ V(c) \}
\]  

(14)
3. Experimental results
In order to ascertain the performance improvement using the IVFKNN classifier scheme, we performed experiments on publicly available Finger Vein USM (FV-USM) image database [7]. The finger vein images were collected at two different sessions in which each subject contributes 24 finger vein images from 4 different fingers: left index, left middle, right index and right middle. Six images were taken from each finger. First session was considered as enrollment or registration and the images from the second session were considered as testing image. The size of captured image was 300 x 100 pixels and 256 grey levels. The benchmark database contains 5904 ROI images from 123 individuals. Principle component analysis (PCA) is employed for feature extraction to generate the most distinguish features [8]. Two experiments were conducted to show the results of the IVFKNN and performance comparisons are made with the competing classifiers: KNN [3] and FKNN [4]. It should be noted that we have specifically chosen to evaluate and compare the performance of the classifiers using classification accuracy, which is one of the most effective measures in the fields of pattern classification. For the first experiment, we intend to evaluate the classification performance between the proposed classifier and the competing classifiers on different sizes of feature dimension of finger vein images. We explore a different numbers of principle component reduction dimensionalities for a fixed image size ratio that is 0.3. Meanwhile, the value of k for each classifier is set equal to 3 in this experiment.

Second experiment aims to evaluate the classification performance between the proposed classifier and the competing classifier over different values of neighborhood size, k. In this experiment, the value of the neighborhood parameter k is preset from 3 to 21 in step 2 while using the optimum values of feature dimension obtained from the previous experiment. It should be stressed out that the optimum value of feature dimensions obtained from the first experiment is used in this experiment. Experiments have been carried out using MATLAB R2014a on Intel(R) Core(TM) i7-4510U CPU processors running @ 2.00GHz and 2.60 GHz. The system has 4 GB of memory and is running a 64-bit Windows 8 operating system.

3.1. Finger Vein USM (FV-USM) Image Database Overview
We have used publicly available finger vein images database [7] that were collected using the proposed acquisition devise explained in [7] from 123 volunteers in USM. The finger vein images were collected at two different sessions in which each subject contributes 24 finger vein images from four different fingers: left index, left middle, right index and right middle. Six images were taken from each finger. First session is considered as enrollment or registration and the images from the second session are considered as testing image. The size of captured image is 300 x 100 pixels and 256 grey levels. The benchmark database contains 5904 ROI images from 123 individuals and it is freely available at the following website: http://drfendi.com for benchmarking/comparison and further/promoting research in this area.

3.2. Classification Performance over Different Numbers of Feature Dimension
As previously stated, the classification performance of the proposed IVFKNN is compared to KNN [3] and FKNN [4] over 2952 finger vein images of USM database by means of classification accuracy. The key objective of this set of experiments was to ascertain the robustness of various algorithms when the finger vein image from both sessions was employed. The parameter neighborhood size, k was fixed as 3 for the first experiment. In the experiment, the size of each image is resized to a fixed resize ratio, 0.3. The original image size is 300 x 100 pixels and with the resize ratio is 0.3, it will be reduced to 0.3 * 300 x 0.3 * 100 pixels. Each classifier is then tested on different numbers of reduction dimensionalities of principle components for each resized image. The experimental results from Table 1 suggests significant improvement in the performance where the proposed IVFKNN classifier achieves the best classification accuracy over the competing classifier. It indicates that the classification of IVFKNN is the most stable than
competing classifiers with varying size of feature vectors. Consequently, the effectiveness of the proposed IVFKNN method with good performance is well manifested on finger vein database in terms of classification accuracy with the corresponding neighborhood size of \( k \).

**Table 1.** The performance of classification accuracy with different number of feature dimensions.

| Feature Dimension | KNN  | FKNN | IVFKNN |
|-------------------|------|------|--------|
| 270               | 76.78| 77.24| 78.12  |
| 540               | 76.86| 77.27| 78.15  |
| 810               | 76.89| 77.30| 78.12  |
| 1080              | 76.89| 77.30| 78.12  |
| 1350              | 76.89| 77.30| 78.12  |
| 1620              | 76.89| 77.30| 78.12  |
| 1890              | 76.89| 77.30| 78.12  |
| 2160              | 76.89| 77.30| 78.12  |
| 2430              | 76.89| 77.30| 78.12  |
| 2700              | 76.89| 77.30| 78.12  |

3.3. Classification Performance over Different Sizes of Neighborhood, \( k \)
In order to prove the robustness of the IVFKNN classifier, the classification performance of IVFKNN as a function of the neighborhood size, \( k \) is investigated on finger vein data set, in comparisons with KNN [3] and FKNN [4]. In this experiment, the optimum values of feature dimension for each classifier were chosen from the previous experiment based on the value that resulting the best classification accuracy. As from the Table 1, the optimum feature dimension for each classifier is 540. The parameter neighborhood size, \( k \) takes the value from 3 to 21 with step 2. The classification rates of different classifiers with varying the neighborhood size, \( k \) are compared to evaluate the sensitivity of the classification performance to the parameter, \( k \). The experimental comparisons in terms of the classification accuracy with varying the neighborhood size, \( k \) are illustrated in Figure 1.

As can be observed in Figure 1, the accuracy rates of IVFKNN is better than other classifiers and stable over the varying number of neighborhood size, \( k \). This fact implies that its classification accuracy is not influenced by the size of neighborhood. The IVFKNN obtain always the highest accuracy for each of the different settings of the \( k \) parameter. Furthermore, when the value of \( k \) is relatively large, the classification differentials between IVFKNN and KNN, FKNN are very significant in most cases. It is noticeable in Figure 1 that the accuracy rates of KNN and FKNN are gradually decreased when the value of \( k \) becomes large. It is due to the fact that their nearest neighbors are dominated by other classes when the neighborhood size increase which leads to misclassification. The experimental results depicted in Figure 1 well indicate that the proposed IVFKNN is the most robust to the neighborhood size \( k \) with promising classification performance, compared to the competing classifiers. The IVFKNN running time is comparable with other competing classifiers as can be observed in Table 2. Other comparisons techniques show a much lower running time but it comes at a cost of a lower accuracy.

4. Conclusion
In this article, a new classification approach for a finger vein recognition called interval valued fuzzy sets \( k \)-nearest neighbors classifier is proposed. With the introduction of intervals, the membership values and neighbor’s voting are represented as an array of intervals. As
Figure 1. The classification accuracy with parameter, $k$ ranging from 3 to 21.

Table 2. The performance comparison in terms of classification accuracy and running time for each classifier with the optimum feature dimension when $k = 3$.

| Classifier | Accuracy(%) | Running time(s) |
|------------|-------------|-----------------|
| KNN        | 76.8970     | 0.0711          |
| FKNN       | 77.3035     | 0.0726          |
| IVFKNN     | 78.1504     | 0.0758          |

consequence, the decision rule can be made in a greater degree of flexibility and thus, improving the classification of IVFKNN than using classical KNN [3] and FKNN [4] classifiers. In order to investigate the classification performance of the proposed classifier, the comprehensive experiments on the database of 2952 finger vein images acquired from 123 subjects were conducted. Two aspects of performance on the proposed classifier are investigated including the classification accuracy performance over different numbers of feature dimension, and the classification accuracy performance with varying the neighborhood size, $k$. The experimental results demonstrated that this new classification rule achieves higher classification accuracy than the conventional KNN [3] and is also superior to the FKNN [4] classification method. In addition, the classification performance of IVFKNN classifier is more robust and stable with varying neighborhood size, $k$.

References
[1] Deng Z et al. 2016 Neurocomputing 195 143
[2] Chen M et al. 2016 Pattern Recogn. 60 486
[3] Cover T and Hart P 1967 IEEE T. Syst. Inform. Theory. 13 21
[4] Keller J M, Gray M R and Givens J A 1985 IEEE T. Syst. Man Cyb. 580
[5] Derrac J et al. 2016 Inform. Sciences 329 144
[6] Bustince H 2000 Int. J. Approx. Reason. 23 137
[7] Asaari M S et al. 2014 Expert Syst. Appl. 41 3367
[8] Wu J D and Liu C T 2011 Expert Syst. Appl. 38 14284