# A GA-based Fuzzy Feature Evaluation Algorithm for Pattern Recognition

Han-Pang Huang* and Yi-Hung Liu**
Robotics Laboratory, Department of Mechanical Engineering
National Taiwan University, Taipei, 10060, Taiwan
TEL/FAX: (886) 2-23633875, E-mail: hphuang@w3.me.ntu.edu.tw
*Professor and correspondence addressee, ** Graduate student

**Abstract**

Feature selection is a very important process in a pattern recognition system. In this paper, a GA-based fuzzy feature evaluation algorithm (GAFEA) is proposed. The relative importance of feature element in the feature space is explored in a 2-D example, and a weighting factor is assigned to each feature element to magnify or reduce its importance in the feature space. In addition, a fuzzy feature evaluation index (*FFEI*) is defined for the measurement of ambiguity of intraclass and interclass. By integrating the weighting factors with the index *FFEI*, the index *FFEI* becomes a function of the weighting set. Based on minimizing *FFEI*, the degrees of the relative importance of features in the multi-dimensional feature space can be found by the genetic algorithm. Finally, an experiment is conducted to justify the validity of the proposed GAFEA. The results show that the task of the feature evaluation can be achieved by the proposed GAFEA before selecting the optimal subset of features.

*Keywords: feature selection, fuzzy feature evaluation, and genetic algorithm.*

## 1. Introduction

Feature selection and classifier design are two key issues in the research of pattern recognition. There are several ways to improve the feature selection. In the past, the statistical technique for feature selection [1] was employed. Fuzzy set approaches to feature selection are based on measures of entropy and the index of fuzziness [2][3]. Pal had successfully found an index of fuzzy evaluation and gave it a good theoretical analysis [4][5]. The recent works for feature selection in the framework of artificial neural network are based mainly on multi-layer feedforward neural networks [6]-[8].

In this paper, a different point of deriving the feature evaluation will be developed in terms of a 2-D example. The relative importance of each feature in the feature set for classifying classes is analyzed qualitatively, and then generalized to an *n*-dimensional feature space. The information of the relative importance of each feature element in the feature space can be expressed in terms of a weighting factor. By incorporating the weighting factors into the feature set, the original feature space can be transformed into a transformed feature space so that the classified results will be higher. Compared with the weighting factors in [4], more precise and physical analysis of the weighting factors is presented in this paper. In addition, a fuzzy feature evaluation index (*FFEI*), which is the modification of the index proposed in [4], is defined. A simple normalized Euclidean norm, rather than a high dimensional norm, is selected to define the distance of a feature set to a center of a class. The size of the class in the proposed index *FFEI* is maintained constant. It will be shown that these simplifications will not influence the resultant ranks in the experiment. Since the index is a function of weighting factors, the task of finding optimal feature subset turns out to minimize the index *FFEI* subject to the weighting set.

Several ways, such as nonlinear programming [9], simulated annealing [10], and genetic algorithm (GA) [11], can be used to perform the optimization task. In papers [4] and [5], the gradient descent algorithm was employed. However, if the relative importance between two or more features is nearly equal, then the gradient descent may lead to a biased rank to a feature due to the property of local minimum of the gradient descent method. In this paper, the genetic algorithm is selected to avoid the result of local minimum since the objective analysis of GA is based on a population of solutions. The proposed GA-based fuzzy feature evaluation algorithm (GAFEA) will experience three stages in the experiment. First, the proposed GAFEA finds out the rank of each feature according to the weighting set. For analyzing objectively, there are 10 runs in the experiment. The second stage is to demonstrate the validity of ranking from the GAFEA by using a *k*-nearest neighbor (*K*-NN) classifier. In this stage, a quantitative analysis is performed. In the final stage, the structure description of classes, i.e., the scatter plot, is used to verify the validity of ranking with a qualitative analysis.

## 2. Relative importance and weighting factors

In this section, the relationship between the relative importance of a feature and the weighting factor will be derived. An example on 2-D feature space is shown in Fig. 1. Assume that there are two classes, $C_1$, and $C_2$ only. The pattern $F_j$ is a *2-D* feature space and $F_j \in C_1$. Distances between $F_j$ and $C_1$, $F_j$ and $C_2$ are calculated with the following Euclidean norm

$$d_1(F_j) = [|F_1 - m_{11}|^2 + |F_2 - m_{12}|^2]^{1/2}$$

$$d_2(F_j) = [|F_1 - m_{21}|^2 + |F_2 - m_{22}|^2]^{1/2}$$

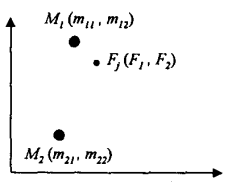



Fig. 1. Example on 2-D feature space.

where $(m_{11},m_{12})$ and $(m_{21},m_{22})$ are, respectively, the centers of class 1 and class 2. Since the pattern $F_j \in C_1$ is known, one can analyze the relative importance of each feature in $F_j$ to discriminate the two classes from geometrical properties in Fig. 1. Suppose that $m_{11} \approx m_{21}$, and the value of $|m_{12} - m_{22}|$ is much larger than $|m_{11} - m_{21}|$. Then $d_2(F_j)$ can be rewritten as: $d_2(F_j) \cong [|F_1 - m_{11}|^2 + |F_2 - m_{22}|^2]^{1/2}$. Since the first terms in $d_1(F_j)$ and $d_2(F_j)$ are quite close, $F_1$ can be viewed as a redundancy for characterizing these classes. Clearly, these two classes can be discriminated by using feature $F_2$ only. Besides, if the distribution of $\{F_1\}_j$, $j$=1,...,$p$, and $F_j \in C_1$, is flat along the $F_1$-axis, i.e., the standard deviation of $\{F_1\}_j$, $j$=1,...,$p$, is large, then the values of $|F_1 - m_{11}|$ and $|F_1 - m_{21}|$ are large. Thus, the distances between the pattern and $M_1$, $M_2$, are primarily controlled by the terms $|F_1 - m_{11}|$ and $|F_1 - m_{21}|$. The weights of the second terms in both $d_1(F_j)$ and $d_2(F_j)$ are so small that one cannot discriminate these two classes by information of the membership values of $F_j$ with respect to classes $C_1$ and $C_2$. This qualitative analysis gives us an idea: for decreasing the weight of the first term ($F_1$) which is useless in characterizing classes, one may add a weighting factor $\omega_1$, $0 < \omega_1 < 1$, such that $\omega_1|F_1 - m_{11}| << |F_1 - m_{11}|$ *and* $\omega_1|F_1 - m_{21}| << |F_1 - m_{21}|$. On the other hand, for the dominant term $F_2$, one may add a weighting factor $\omega_2$, $0 < \omega_2 < 1$, $\omega_2 > \omega_1$, such that $\omega_2|F_2 - m_{12}| > \omega_1|F_1 - m_{11}|$ *and* $\omega_2|F_2 - m_{22}| > \omega_1|F_1 - m_{21}|$ As a matter of fact, the inductive results of the 2-dimensional example can be generalized to an *n-dimensional* feature space.

### 3. Fuzzy feature evaluation index

Consider a pattern $F$ with an $n$-dimensional feature space $F$=($F_1$, $F_2$, ..., $F_i$, .., $F_n$). Suppose that there are $l$ classes $C_1$,..., $C_k$,....., $C_l$, and there are $p$ pattern samples in each class, where $p$ is the class size. The distance between the pattern $F$ and the center of the class is defined by the normalized Euclidean norm as

$$d_k(F) = [\sum_{i=1}^{n} (\frac{F_i - m_{ki}}{\alpha_{ki}})^2]^{1/2} \quad (1)$$

where

$$m_{ki} = \frac{1}{p}\sum_{j=1}^{p} F_{ij} \quad (2)$$

and

$$\alpha_{ki} = N \cdot \max_j |F_i - m_{ki}| \quad (3)$$

Note that $m_{ki}$ is the center (mean) of the *i-th* feature of $F$ in class $k$ and $F \in C_k$. Due to Eq. (3), $d_k(F)$ in Eq. (1) can be normalized to [0,1] and $N$ is a normalized factor. The membership function of a pattern $F$ to a specific class $C_k$ is defined as

$$\mu_k(F) = \begin{cases} 1 - 2d_k^2(F), & \text{if } 0 \le d_k(F) < 1/2 \\ = 2[1 - d_k(F)]^2, & \text{if } 1/2 \le d_k(F) \le 1 \\ = 0, & \text{otherwise.} \end{cases} \quad (4)$$

The fuzzy feature evaluation index (*FFEI*) is defined as

$$FFEI = \sum_{k=1}^{l} \sum_{F \in C_k} \frac{E_k(F)}{\sum_{k' \neq k} E_{kk'}(F)} \times \frac{1}{p} \quad (5)$$

where

$$E_k(F) = \mu_k(F) \times (1 - \mu_k(F)) \quad (6)$$

and

$$E_{kk'}(F) = [\mu_k(F) \times (1 - \mu_{k'}(F))] + [\mu_{k'}(F) \times (1 - \mu_k(F))] \quad (7)$$

Note that $\mu_k(F)$ and $\mu_{k'}(F)$ are membership values of the pattern $F$ in class $C_k$ and $C_{k'}$. The index $E_k(F)$ represents the compactness of an individual class, i.e., the ambiguity measurement of the intraclass. From Eq. (6), one can see that $E_k(F)$ is minimum (zero) when $\mu_k(F)$=1 or 0 and maximum (0.25) when $\mu_k(F)$=0.5. The other index $E_{kk'}(F)$ represents the separation between classes $\forall k \neq k'$, i.e., the ambiguity measurement of interclass. The index $E_{kk'}(F)$ is minimum (zero) when $\mu_k(F) = \mu_{k'}(F)$ =1 or $\mu_k(F) = \mu_{k'}(F)$ =0 and is maximum when $\mu_k(F)$ =1, $\mu_{k'}(F)$ =0 or $\mu_k(F)$ =0, $\mu_{k'}(F)$ =1. The term $E_k(F)/\sum E_{kk'}(F)$ is minimum when $\mu_k(F)$ =1 and $\mu_{k'}(F)$=0 and is maximum when $\mu_k(F)$ =0.5 for all $k$. The index *FFEI* is a compound of $E_k(F)$ over $E_{kk'}(F)$. The qualitative analysis of the index *FEEI* is described as follows. The value of *FEEI* decreases when the pattern $F$ increases its membership value to only one class, i.e., $F$ increases its belongingness to only one class, and in the meantime, $F$ decreases its ambiguity of interclass. On the other hand, the value of *FEEI* increases when the membership value of the pattern $F$ for every class approaches 0.5, i.e., $F$ decreases its belongingness to a specific class $C_k$, and in the meantime, it increases the ambiguity to other classes for some $k' \neq k$. Hence, the feature selection becomes the task of the minimization of the index *FFEI*, i.e., one can find the optimal subset of feature space $F$ via minimizing the index *FFEI*. However, there are no arguments in Eqns. (1)-(7) for which we can minimize the index *FFEI* subject to them. Namely, there is no information about the relative importance of each feature component $F_i$ in the n-dimensional feature space. Therefore, the weighting factors mentioned in the previous section are incorporated into the normalized Euclidean distance in Eq. (1) as

$$d_k(F) = [\sum_{i=1}^{n} (\omega_i \cdot \frac{|F_i - m_{ki}|}{\alpha_{ki}})^2]^{1/2} \quad (8)$$

where $\omega_i$, $0 < \omega_i < 1$ for all $1 \le i \le n$, is the weighting factor of the *i-th* feature in the $n$-dimensional feature space. It also reflects the relative importance of the *i-th* feature to the others in characterizing classes. The higher of the value of $\omega_i$ is, the greater of the importance of the *i-th* feature is. Let $\varpi = (\omega_1, \omega_2, ...., \omega_n)$ be the weighting set. Now the index *FFEI* is a function of the set $\varpi$. After finding the set $\varpi$, each important degree of features can be found and one can select subset of salient features from

the $n$-dimensional feature set. Thus, the feature selection becomes the task of optimizing index *FFEI* subject to the set $\varpi$.

## 4. Minimization of the *FFEI* via GA

The gradient descent algorithm is frequently used to solve the problem with an analytic solution, but a local minimum and a complex process of deviation are often followed. For avoiding falling into a local minimum, the genetic algorithm is a good choice as long as the genetic operations, such as population size and mutation rate, are properly determined. The genetic algorithm searches for a population of solutions rather than a single solution. This differs from conventional optimization methods, and provides the merit of global optimum. Another reason of choosing the genetic algorithm is that it is working under a process of simulation since the feature selection is an off-line process. The only weakness of using GA is long convergent time. In fact, this weakness can be overcome by using a high-speed personal computer.

## 5. Experiment results

After deriving the GAFEA, a well-known data, Anderson's Iris data [12], is used to validate the algorithm. Anderson's Iris data contains three classes: *Iris setosa* (class 1), *Iris versicolor* (class 2), and *Iris virginica* (class 3). Each class contains 50 samples, and each sample has four features: *sepal length* (*SL*), *sepal width* (*SW*), *petal length* (*PL*), and *petal width* (*PW*). Before performing GA to minimize the index *FFEI*, some genetic operations in GA search must be set first. All weighting factors are randomized in [0,1] initially, and then encoded to binary 10-bit strings (chromosomes). Therefore, the resolution of a weighted factor is 1024 in [0,1] such that the relative importance can be obtained more precisely. Besides, the population size is set to 20, the mutation rate is set to 0.01%, and the crossover rate is set to 1.0. A roulette wheel selection is used as the evolution operation. The fitness function is defined as the inverse of the index *FFEI*, and the cost of each generation is defined as the summation of all populations in the generation. A genetic algorithm search with 50 generations is called a run. There are 10 runs in our experiment such that the results can be compared objectively. Relative importance and rank of the 10 runs are listed in Table 1. It is worth mentioning that the time of each run is less than one minute by using a personal computer with PIII-450 CPU. These weighting factors are determined when the convergence curve keeps at a constant. The rank of a feature is determined by its corresponding weighting factor, $\omega$, i.e., the relative importance. The higher the rank is, the more important the feature is. The ranks of feature *PL* and *PW* are the first or the second in the four features, and the ranks of feature *SL* and *SW* are the last two. Furthermore, there exists a great difference of order between the pair {*PL, PW*} and the pair {*SL, SW*}. It indicates that the pair {*PL, PW*} is much more important than the pair {*SL, SW*}. However, the order between the features *PL* and *PW* is almost the same. It indicates that the importance of features *PL* and *PW* are nearly equal, and so are those of *SL* and *SW*. So far, we have found out the feature subset {*PL, PW*} with higher ranking, and the feature subset {*SL, SW*} with lower ranking by the GAFEA. In order to demonstrate the validity of the ranking from the GAFEA, the $k$-nearest neighbor ($K$-NN) algorithm is used as the classifier to obtain the classification rates for all permutations of pairs in the four features. For avoiding the existence of a tie [13], the $k$ is set as an odd value, $k$=3. The classification results are listed in Table 2. From the results, it is found for the pair {*PL, PW*} that the $K$-NN classifier results in 100%, 98%, 96% for the three classes, and the average classification rate is 98%. It is much higher than the results given by the pair {*SL, SW*}. In particular, for all features {*SL, SW, PL, PW*}, each classification rate and the average classification rate are smaller than the results using the pair {*PL, PW*} only. Obviously, the pair {*SL, SW*} will encumber the classification rate when selecting all the features to classify the three classes. Another way to demonstrate the validity of the ranking is from the point of the structural description of classes. The scatter plots of all pairs of features are shown in Fig. 2(a)-(f). The symbols ".", "+", and "*" represent the classes *Iris setosa* (class 1), *Iris versicolor* (class 2), and *Iris virginica* (class 3), respectively. The distribution of class 1 structure in each scatter plot is far away from others. The ambiguity between class 2 and class 3 is the largest in the scatter plot of *SL-SW*, but is the smallest in the scatter plot of *PL-PW*. Again, these scatter plots show the validity of the ranking resulted from the GAFEA.

## 6. Conclusions

In this paper, a GA-based fuzzy feature evaluation algorithm (GAFEA) has been proposed. A fuzzy feature evaluation index (*FFEI*) is used to measure the compactness of intraclass and the ambiguity of interclass. However, the major contribution is to explore the weighting factors and relative importance of features. The weighting factor is integrated into the index *FFEI* so that the index is a function of the factor. Then the genetic algorithm is performed to determine the global optimal set of the weighting factors. The experimental results show that the proposed GAFEA can not only find out the optimal feature subset but also accomplish the task of feature evaluation in a short time.

## References


1. P. A. Devijver and J. Kittler. *Pattern Recognition: A Statistical Approach.* Prentice-Hall International, 1982.
2. S. K. Pal and Basabi Chakraborty, "Intraclass and interclass ambiguities (fuzziness) in feature evaluation," *Pattern Recognition Latter*, vol. 2, pp. 275-279, 1984.
3. S. K. Pal and Basabi Chakraborty, "Fuzzy set theoretic measure for automatic feature evaluation," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 16, pp. 754-760, 1986.
4. S. K. Pal, J. Basak, and R. K. De, "Fuzzy feature evaluation index and connectionist realization," *Information Science*, vol. 105, pp173-188, 1998.

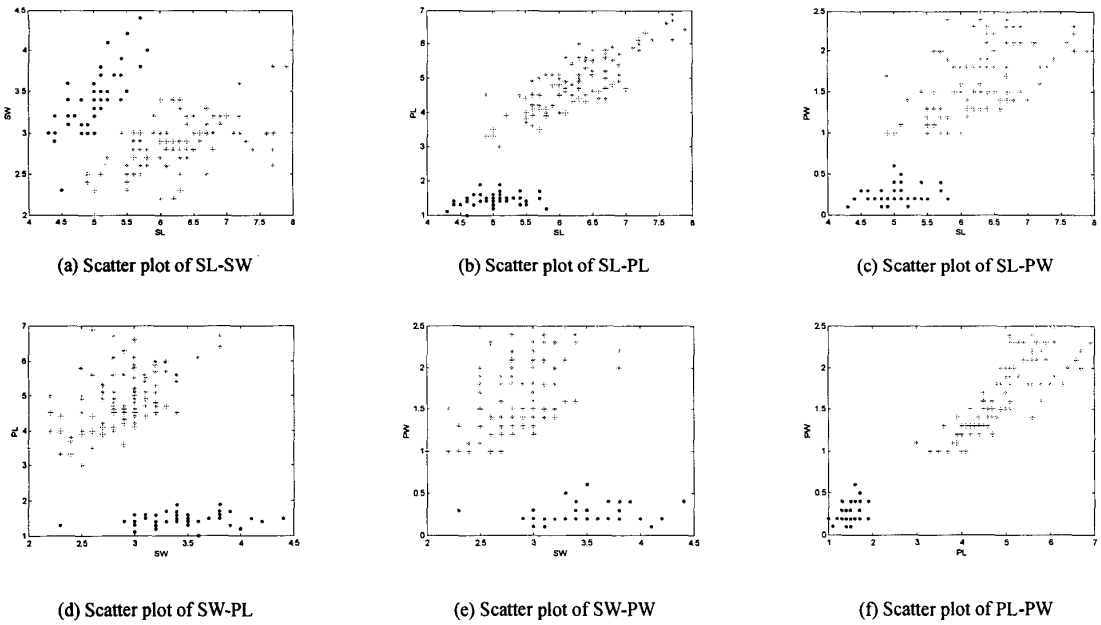
5. J. Basak, R. K. De, and S. K. Pal, "Fuzzy feature evaluation index and connectionist realization-II: Theoretical analysis," *Information Science*, vol. 111, pp1-17, 1998.
6. K. L. Priddy, S. K. Rogers, D. W. Ruck, G. L. Tarr, and M. Kabrisky, "Bayesian selection of important features for feedforward neural network," *Neurocomputing*, vol. 5, pp. 91-103, 1993.
7. W. A. C. Schmidt and J. P. Davis, "Pattern recognition properties of various feature spaces for higher order neural networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.* 15, pp. 795-801, 1993.
8. R. K. De, N. R. Pal, and S. K. Pal, "Feature analysis: Neural network and fuzzy set theoretic approaches," *Pattern Recognition*, vol. 30, pp. 1579-1590, 1997.
9. D. M. Himmelblau, *Applied Nonlinear Programming*, New York: McGrawHill, 1976.
10. Mitsuo Gen, and Runwei Cheng, *Genetic Algorithms & Engineering Design*, John Wiley & Sons, Inc. 1997.
11. P. J. M. van Laarhoven, and E. H. L. Aarts, *Simulated Annealing: Theory and Applications*, Published by D. Reidel Publishing Company, 1987.
12. R. A. Fisher, "The use of multiple measurements in taxonomic problem," *Annals of Eugenics*, vol. 7, pp. 179-188, 1936.
13. James M. Killer, Michael R. Gray, and James A. Givens, JR., "A fuzzy k-nearest neighbor algorithm," *IEEE Transactions on Systems, Man, and Cybernetics,* vol. 15, no. 4, pp. 580-585, 1986.


Table 1. Relative importance and rank of each feature in each run

| | Initial ω in [0,1] | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | | 4 | | 5 | | 6 | | 7 | | 8 | | 9 | | 10 | |
| | ω | *rank* | ω | *rank* | ω | *rank* | ω | *rank* | ω | *rank* | ω | *rank* | ω | *rank* | ω | *rank* | ω | *rank* | ω | *rank* |
| *SL* | 0.0645 | 4 | 0.0391 | 3 | 0.0752 | 4 | 0.0273 | 4 | 0.1484 | 4 | 0.0410 | 3 | 0.0947 | 3 | 0.0381 | 4 | 0.0488 | 4 | 0.0566 | 4 |
| *SW* | 0.1260 | 3 | 0.0254 | 4 | 0.1123 | 3 | 0.0381 | 3 | 0.1641 | 3 | 0.0371 | 4 | 0.0850 | 4 | 0.1270 | 3 | 0.3574 | 3 | 0.0576 | 3 |
| *PL* | 0.5703 | 1 | 0.3877 | 2 | 0.1953 | 2 | 0.2295 | 2 | 0.5479 | 1 | 0.4541 | 1 | 0.8701 | 1 | 0.6338 | 1 | 0.7979 | 1 | 0.4902 | 1 |
| *PW* | 0.2432 | 2 | 0.4482 | 1 | 0.3857 | 1 | 0.4688 | 1 | 0.3037 | 2 | 0.3359 | 2 | 0.6641 | 2 | 0.4072 | 2 | 0.6689 | 2 | 0.2354 | 2 |

Table 2. Comparison of classification rates among different pairs of features of iris data by *K*-NN classifier

| | {*SL, SW*} | {*SL, PL*} | {*SL, PW*} | {*SW, PL*} | {*SW,PW*} | {*PL, PW*} | {*SL, SW, PL, PW*} |
|---|---|---|---|---|---|---|---|
| *Class 1 (setosa)* | 98 % | 100 % | 100 % | 100 % | 100 % | 100 % | 100 % |
| *Class 2(versicolor)* | 80 % | 96 % | 94 % | 94 % | 98 % | 98 % | 94 % |
| *Class 3 (virginica)* | 76 % | 92 % | 94 % | 94 % | 92 % | 96 % | 94 % |
| *Average* | 84.67 % | 96 % | 94.67 % | 94.67 % | 96.67 % | 98 % | 96 % |




(a) Scatter plot of SL-SW (b) Scatter plot of SL-PL (c) Scatter plot of SL-PW

(d) Scatter plot of SW-PL (e) Scatter plot of SW-PW (f) Scatter plot of PL-PW

Fig. 2. Scatter plot of each pair of features for Iris data